



# Deep learning robotic guidance for autonomous vascular access

Alvin I. Chen<sup>ID</sup>✉, Max L. Balter<sup>ID</sup>, Timothy J. Maguire and Martin L. Yarmush

**Medical robots have demonstrated the ability to manipulate percutaneous instruments into soft tissue anatomy while working beyond the limits of human perception and dexterity. Robotic technologies further offer the promise of autonomy in carrying out critical tasks with minimal supervision when resources are limited. Here, we present a portable robotic device capable of introducing needles and catheters into deformable tissues such as blood vessels to draw blood or deliver fluids autonomously. Robotic cannulation is driven by predictions from a series of deep convolutional neural networks that encode spatiotemporal information from multimodal image sequences to guide real-time servoing. We demonstrate, through imaging and robotic tracking studies in volunteers, the ability of the device to segment, classify, localize and track peripheral vessels in the presence of anatomical variability and motion. We then evaluate robotic performance in phantom and animal models of difficult vascular access and show that the device can improve success rates and procedure times compared to manual cannulations by trained operators, particularly in challenging physiological conditions. These results suggest the potential for autonomous systems to outperform humans on complex visuomotor tasks, and demonstrate a step in the translation of such capabilities into clinical use.**

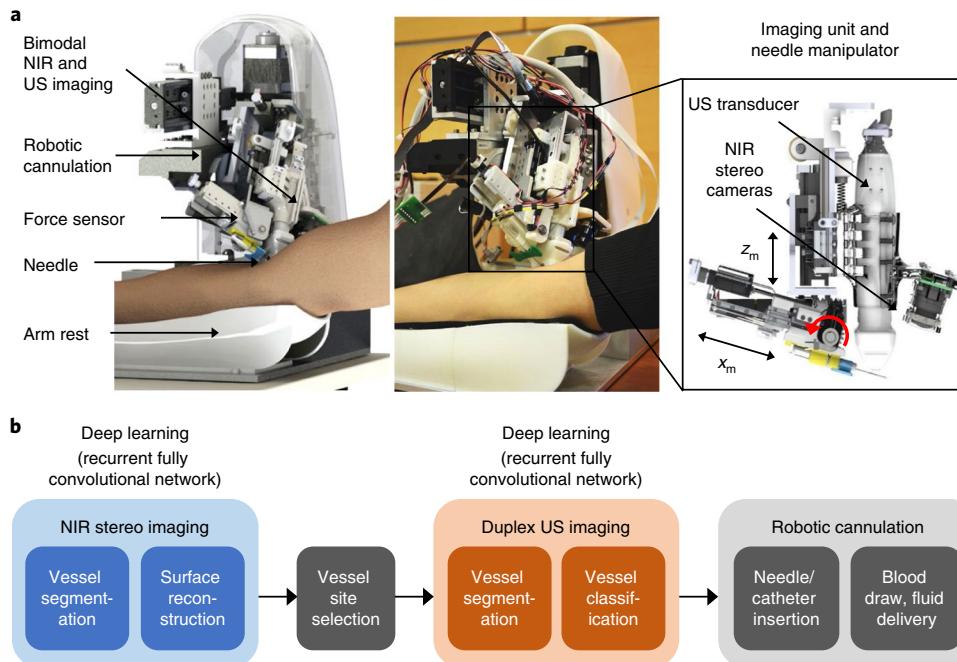
Medical robots promise enhanced precision, safety and efficacy by working beyond the limits of human perception and dexterity<sup>1,2</sup>. Recent advancements in image guidance, perception, robotic planning and computing have provided medical robotic systems with the intelligence and manoeuvrability to perform challenging interventional tasks on soft tissues with supervised autonomy and comparable performance to trained practitioners<sup>3–7</sup>. One task in which positioning guidance is particularly important is percutaneous cannulation of soft tissues such as blood vessels. Gaining access to vessels is a critical first step in a plethora of diagnostic and therapeutic procedures, including drawing blood, administering fluids and medications, introducing endovascular devices and monitoring physiological status<sup>8,9</sup>. The timely delivery of these interventions can affect morbidity and mortality<sup>10</sup>, yet, in difficult conditions, vascular access can be highly challenging. In remote and resource-limited environments, medical personnel are often required to perform life-saving tasks under the most chaotic circumstances<sup>11</sup>. Failures are estimated to occur in 20% of procedures, and difficulties are exacerbated in patients with small (<1 mm diameter), tortuous or collapsed vessels, which are common in paediatric, elderly, chronically ill and traumatic patients<sup>12,13</sup>. In these groups, first-stick accuracies fall below 50% and five or more cannulation attempts are commonly needed, leading to delays in access and treatment<sup>14</sup>. Major bleeding complications can arise when critical adjacent tissues (major arteries, nerves or internal organs) are punctured, and the risk of complication increases significantly with multiple cannulation attempts<sup>15</sup>. When peripheral vessels are inaccessible, more invasive approaches such as central venous or arterial access are often required<sup>9</sup>.

The challenges of difficult vascular access have driven the development of imaging technologies that fall into four main categories: (1) tactile pressure-based imaging, which can provide maps of tissue elastic response with sensitivities of several pascals but at poor spatial resolutions (>1 mm)<sup>16</sup>; (2) optical coherence tomography and photoacoustic tomography, which have demonstrated spatial resolutions of 0.01–0.1 mm but with limited imaging depth (1–2 mm)<sup>17–20</sup>; (3) near-infrared (NIR) optical imaging, which

utilizes 700–1,000 nm light from lasers or light-emitting diodes (LEDs) to image superficial vessels within 5 mm of the tissue surface<sup>21,22</sup>, but which does not accurately estimate vessel depth beneath the skin; (4) ultrasound (US) imaging, which has seen the greatest clinical adoption and been correlated to higher vascular access success and lower complication rates compared to blind cannulation<sup>23–25</sup>. Modern clinical linear-array US transducers can resolve submillimetre tissue structures at suitable depths (0.5–10 cm for frequencies of 5–20 MHz) and estimate blood flow velocities using Doppler-based modalities. The main limitation of US is the expertise required to obtain optimal views of thin structures in three-dimensional (3D) space based on 2D planar images. As a result, US-guided vascular access is most commonly performed by clinical personnel with specialized training<sup>26</sup>.

Unlike imaging-based methods, which rely on manual insertion, robotic strategies could altogether eliminate the dependence on practitioner experience and availability<sup>1–3</sup>. Commercial robotic catheter systems have been approved by the United States Food and Drug Administration for navigating inside peripheral vessels during endovascular therapy<sup>27</sup>, although these systems are not image-guided and only help with intravascular navigation following successful access. Force, tactile and impedance sensing have been reported for detecting vessel puncture events during robotic needle insertion, although the initial steps of needle positioning and steering are still performed manually<sup>28–31</sup>. More recently, robotic systems for vessel cannulation have been investigated utilizing duplex (simultaneous B-mode and colour Doppler) US vessel imaging<sup>32,33</sup>, monocular NIR imaging<sup>34</sup>, NIR stereo imaging<sup>35,36</sup> or multimodal imaging<sup>37,38</sup>, but closed-loop guidance of these systems has not been demonstrated and the feasibility of autonomous operation has not yet been achieved.

Here, we present a portable robotic device capable of steering needles and catheters into submillimetre vessels with minimal supervision (Fig. 1 and Supplementary Fig. 1). Autonomous robotic guidance is driven by a deep learning<sup>39</sup> framework that takes bimodal NIR and duplex US imaging sequences as its inputs and performs a series of complex vision tasks, including vessel



**Fig. 1 | Autonomous image-guided robotic vascular access, blood drawing and fluid delivery.** **a**, The device couples bimodal 3D vessel imaging and adaptive robotic instrument manipulation within a compact shell. The robot comprises a six-degrees-of-freedom (6-DOF) base positioning unit and a 3-DOF distal manipulator that performs autonomous cannulation under NIR and US image guidance. **b**, End-to-end workflow for robotic vascular access. The robotic guidance is driven by a pair of trained deep convolutional neural networks operating on continuous NIR and US imaging sequences.

segmentation, classification and depth estimation. Using the device, we evaluate image-guided robotic tracking in humans, and we compare autonomous robotic cannulation to manual performance *in vitro* (in tissue-mimicking phantoms simulating broad demographic variability) and *in vivo* (anaesthetized rat models of superficial venous access).

## Results

**End-to-end workflow for robotic vascular access.** The device relies on the complementary use of NIR and US imaging (Fig. 2a) to achieve an end-to-end robotic workflow. NIR imaging provides non-contact visualization of superficial vessels over a broad ( $20 \times 15$  cm) field of view, while US imaging allows focal visualization of a target vessel and facilitates submillimetre pose adjustments to compensate for vessel motion. The robotic cannulation involves a sequence of automated tasks:

1. Automatically load, and later dispose of the needle.
2. Scan the arm under NIR stereo imaging and reconstruct a 3D surface map of segmented vessels.
3. Identify a suitable vessel as the target cannulation site (currently performed by the operator).
4. Robotically position the US transducer above the target vessel while compensating for arm motion.
5. Segment and track the target vessel in the US image while differentiating arteries from veins.
6. Robotically align the needle with the target vessel in 6-DOF space while compensating for arm motion.
7. Guide the needle into the target vessel under US image feedback, again while compensating for vessel motion.
8. Confirm successful lumen access or identify failed cannulation based on US and force feedback.
9. Draw blood or deliver fluids.

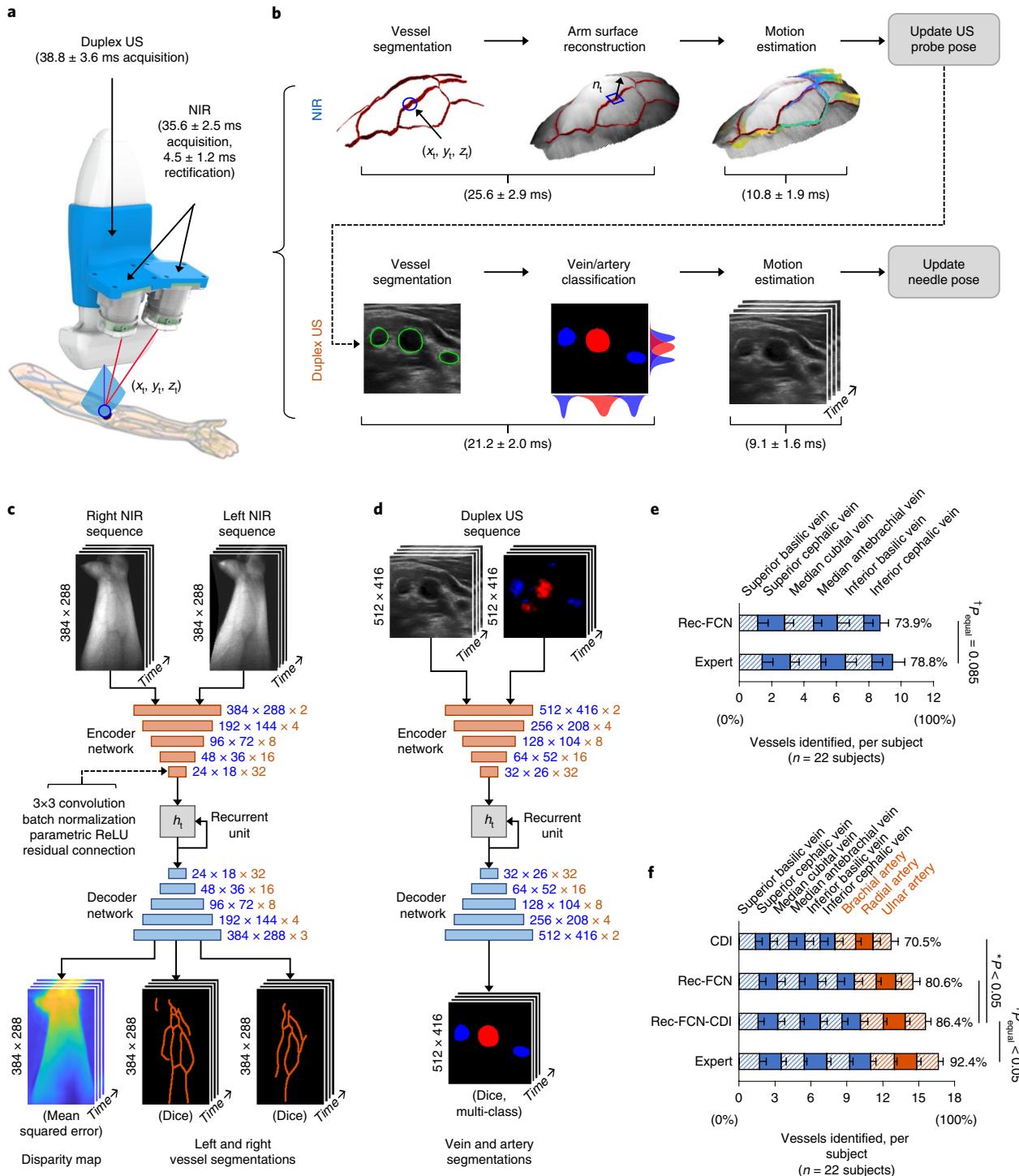
The device automates the handling of disposables before and after the procedure (Supplementary Fig. 2a) and is designed to be

able to instantaneously release the needle on detection of sudden motions (Fig. 4e,f) or excess forces at the needle tip (Supplementary Fig. 2b-e) during cannulation. The device is further capable of drawing blood into sample collection vials or advancing peripheral catheters up to 25 mm beyond the access point (Supplementary Fig. 1b).

**Deep learning encodes spatiotemporal information for autonomous image guidance.** In the NIR guidance step, the robot infers from a deep neural network trained to simultaneously segment, track and compute the depth of peripheral vessels from stereo camera image sequences (Fig. 2b, top). The deep learning model is based on a recurrent fully convolutional network (Rec-FCN) architecture<sup>40,41</sup> whose design attempts to capture salient image features and motion signatures at multiple resolution scales (Fig. 2c and Supplementary Figs. 3–5). The network takes, as input, stereo image pairs from the current frame and encoded features generated by the network in previous frames. The outputs are a pair of dense vessel segmentations and a dense disparity map (Supplementary Fig. 3). We derive a 3D map of the arm surface and vasculature, from which the operator may select a target vessel that is subsequently tracked in real time in the presence of arm motion.

In the US guidance step, the robot positions the transducer based on the 3D vessel pose computed according to the predictions of the first network and lowers the transducer against the arm surface (Fig. 2b, bottom). A second network, also based on the Rec-FCN architecture, operates on B-mode and colour Doppler image (CDI) sequences to predict dense segmentations of veins and arteries (Fig. 2d and Supplementary Fig. 4).

**Automated vessel segmentation and stereo reconstruction from NIR image sequences.** We compared expert manual vessel segmentation and deep learning segmentation from NIR stereo image sequences in left and right forearms of 22 adult volunteers ( $n=44$ ). Deep learning segmentation based on the Rec-FCN architecture



**Fig. 2 | Bimodal vessel imaging and analysis with deep convolutional neural networks.** **a,b**, NIR and US image acquisition (**a**) and analysis (**b**) along with processing times for each step (for details see Supplementary Fig. 11). The point  $(x_t, y_t, z_t)$  indicates the cannulation target selected based on NIR imaging. Vector  $n_t$  indicates the normal direction from the arm surface at  $(x_t, y_t, z_t)$ , computed from NIR stereo reconstruction. **c**, Recurrent fully convolutional network (Rec-FCN) operating on NIR image sequences. Inputs are two-channel tensors (dimensions 384 × 288 × 2) comprising rectified left and right stereo image pairs. Outputs are three-channel tensors (384 × 288 × 3) comprising left vessel segmentations, right vessel segmentations and disparity predictions, from which 3D vessel orientation is computed.  $h_t$  indicates the network's encoded features. **d**, Rec-FCN model operating on B-mode and colour Doppler image (CDI) sequences. Inputs are two-channel tensors (dimensions 512 × 416 × 2). Outputs are two-channel tensors (512 × 416 × 2) of vein and artery segmentations. **e**, Comparison of upper-extremity veins identified under NIR imaging by Rec-FCN and expert assessment. Positive detection is defined as mean Dice score exceeding 0.70 across all image frames containing the target vessel. **f**, Comparison of upper-extremity veins and arteries identified by manual segmentation under US guidance, clutter filtering of CDI sequences, Rec-FCN segmentation of B-mode sequences and segmentation of concatenated B-mode and CDI sequences (Rec-FCN-CDI). Positive detection is defined as mean Dice score exceeding 0.70 across all image frames containing the target vessel (**f**). Bars show mean ± s.d. of number of vessels identified across participants. \*Equivalence determined using a two one-sided t-tests (TOST) procedure (**e,f**). \*Kruskal-Wallis one-way analysis of variance (ANOVA) with Dunn's post hoc test (**f**).

was able to detect a majority of upper-extremity venous branches that were observed by manual assessment (73.9% (195/264) of veins detected by Rec-FCN compared to 78.8% (208/264) by manual assessment; two one-sided *t*-test (TOST) equivalence test,  $P=0.085$ ; positive detection defined as mean Dice score  $>0.70$  across all image frames containing the vessel) (Fig. 2e). We also compared the pixel-wise accuracy of the predictions to expert annotations based on Dice score ( $0.80 \pm 0.18$ ), Jaccard index ( $0.74 \pm 0.20$ ) and modified Hausdorff distance ( $2.06 \pm 1.69$  mm overall,  $1.09 \pm 0.85$  at the cannulation target) (Fig. 3a,b and Supplementary Method 1). Inference on image sequences improved the segmentation accuracy compared to fully convolutional networks (FCNs) without recurrence operating on independent image frames (Fig. 3a).

Three-dimensional arm surface reconstructions computed from predicted disparity maps (Supplementary Fig. 6) had magnitude errors (calculated based on modified Hausdorff distance) of  $2.52 \pm 0.87$  mm compared to ground-truth measurements ( $2.24 \pm 0.75$  mm error at the cannulation site) (Fig. 3a), which was sufficiently accurate to guide robotic placement of the US transducer over the skin. Again, the recurrent network produced lower reconstruction errors compared to single-frame predictions. Finally, features extracted from the segmentations were predictive of the suitability of individual branches for cannulation (Supplementary Fig. 8).

**Automated vessel segmentation and classification from duplex US sequences.** In the subsequent US visualization, we compared manual and deep learning vessel segmentation and classification in 22 volunteers. Rec-FCN segmentation based on concatenated two-channel duplex images detected the majority of venous and arterial branches in the forearm that were identified by manual assessment (86.4% (342/396) vessels detected compared to 92.4% (366/396) by manual assessment; TOST equivalence test,  $P=0.036$ ; positive detection defined as mean Dice score  $>0.70$  across all image frames containing the vessel; Fig. 2f). Furthermore, the method demonstrated significantly increased sensitivity compared to Rec-FCN segmentation based on single-channel B-mode images alone (319 of 396 (80.6%) vessels detected; 87.2% sensitivity; Kruskal–Wallis one-way ANOVA with Dunn's post hoc test,  $P=0.025$ ) as well as to clutter filtering<sup>42</sup> of the CDI alone (279 of 396 (70.5%) vessels detected; 76.2% sensitivity;  $P<0.0001$ ). We observed submillimetre segmentation errors relative to expert annotations (modified Hausdorff distance  $0.54 \pm 0.64$  mm; Dice score  $0.84 \pm 0.12$ ; Jaccard index  $0.77 \pm 0.21$ ) and saw that predictions on sequences outperformed predictions on single frames (Fig. 3a,c and Supplementary Fig. 9).

Analysis of statistical measures of binary classification (Supplementary Method 2) revealed that Rec-FCN predictions from either B-mode or concatenated two-channel duplex images were more reliable in differentiating veins and arteries than CDI clutter filtering<sup>42</sup> (Fig. 3d). ROC curves (Fig. 3e), precision-recall curves (Fig. 3f) and individual classification metrics (Supplementary Fig. 10) showed that inference on temporal sequences again improved classification performance compared to single-frame predictions.

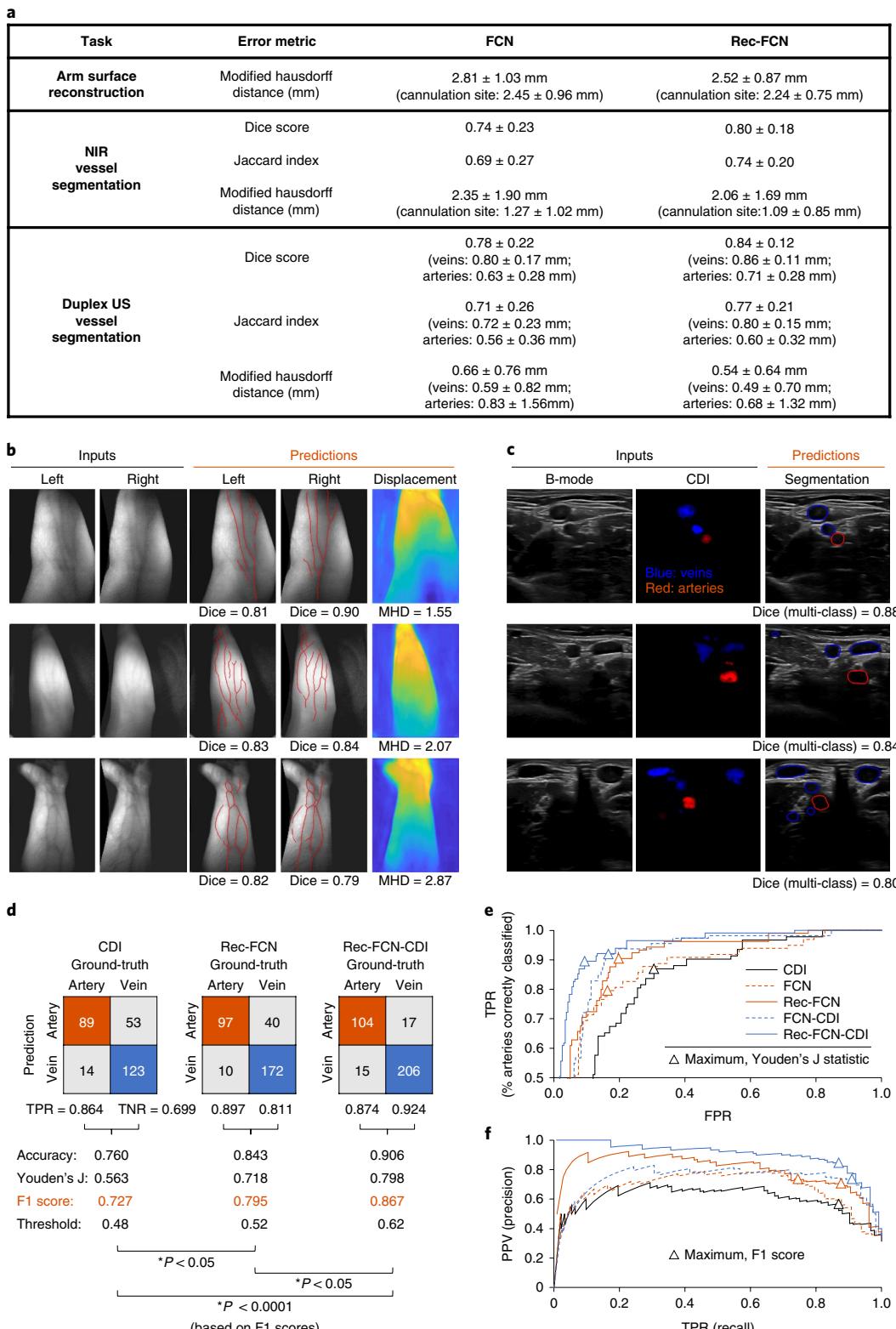
**Real-time robotic tracking and motion compensation.** During NIR-guided tracking, the device compensates for arm motion by continuously computing the 3D pose of segmented vessel targets and making submillimetre adjustments to keep the target within the centre of the imaging field of view (FOV). Similarly, after the US transducer is lowered, the robot adjusts the pose of the US transducer to maintain skin contact and vessel alignment. We evaluated NIR- and US-guided robotic tracking under widely varying arm motions in left and right forearms of 13 adult volunteers (Fig. 4a,b). Tracking errors resulting from Rec-FCN predictions

on image sequences ( $1.65 \pm 1.07$  mm and  $1.82 \pm 1.85^\circ$  mean absolute translational and rotational errors under NIR guidance;  $0.97 \pm 0.25$  mm and  $1.55 \pm 1.40^\circ$  under US guidance) were found to be lower than errors based on predictions from single image frames (Supplementary Fig. 12).

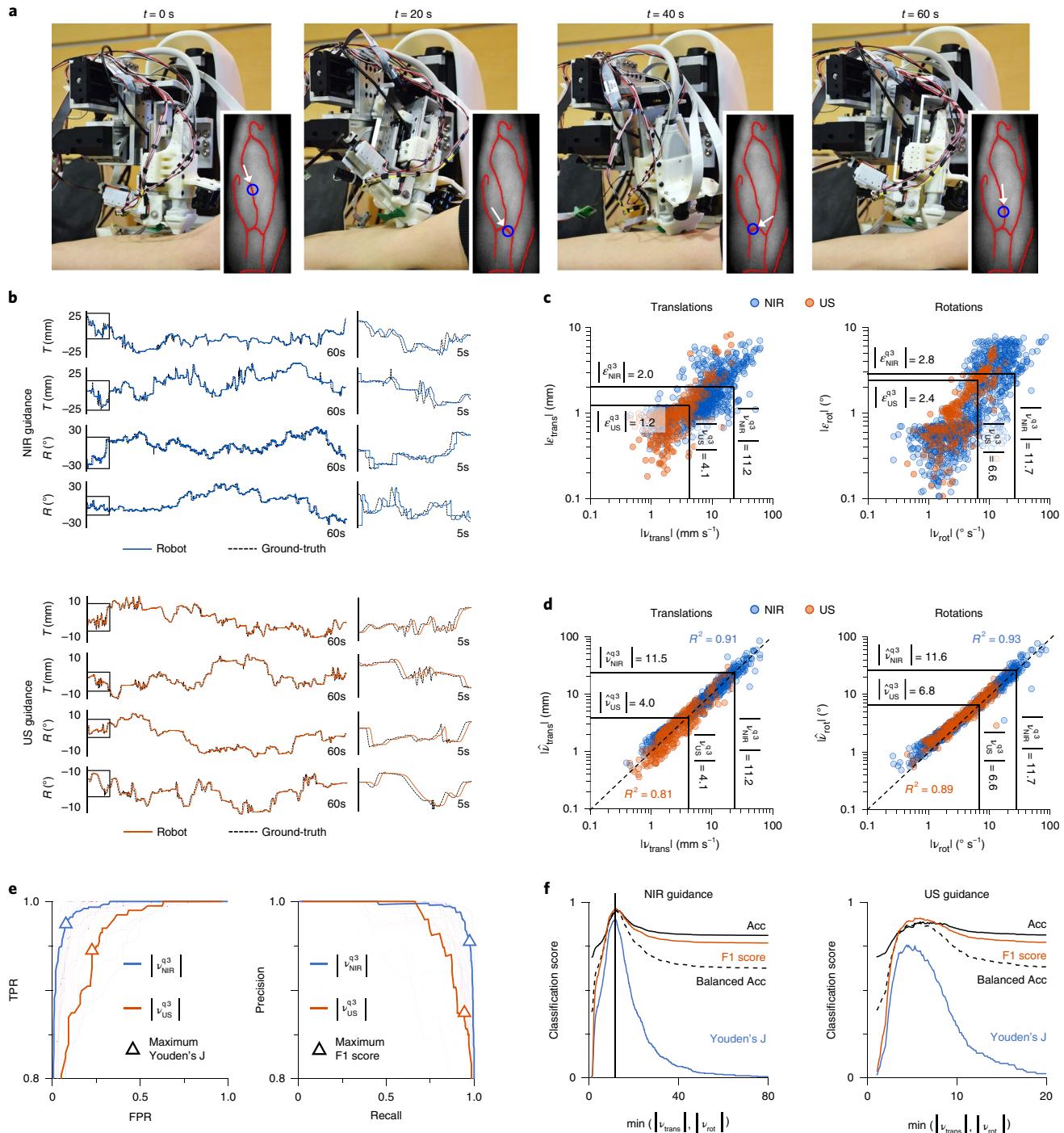
We observed that faster arm motions resulted in larger tracking errors (Fig. 4c). In particular, sudden translations and rotations that exceeded the device's velocity limits would lead to a temporary lag in the tracking trajectory (Fig. 4b). We investigated whether the device could detect these sudden changes as a potential way to minimize injury risk; retrospective analysis of the NIR and US sequences showed that frame-to-frame motion estimates derived from the predicted segmentations (Supplementary Fig. 7) correlated strongly with true frame-to-frame displacements (based on manual segmentations) across the range of motions seen in the study (Fig. 4d). Furthermore, we found that the frame-to-frame estimates could serve as reliable indicators of sudden motion (Fig. 4e,f) and thereby facilitate the deployment of critical safety mechanisms in an automated manner.

**In vitro autonomous cannulation across a broad physiological spectrum.** We next investigated the effects of physiological and demographic variability on robotic performance using multilayered tissue-mimicking phantom models. The models comprised five tissue layers (epidermis, dermis, hypodermis, blood vessels and blood) and were tailored to reproduce the mechanical, optical and acoustic properties of human tissues over a broad demographic range (Fig. 5a and Supplementary Fig. 13)<sup>43</sup>. We applied a fractional factorial experimental design to optimize robotic insertion speed, insertion angle and needle size (Supplementary Fig. 14a) and assess device performance (using the optimized insertion settings) across 15 tissue properties (Supplementary Fig. 14c). We found that the insertion parameters that maximized robotic performance were not constant but instead varied across different tissue conditions (Supplementary Fig. 14b). Specifically, four of the 15 tissue properties (skin tone, hypodermis thickness, hypodermis elasticity and vessel diameter) were found to significantly influence cannulation success (Fig. 5b and Supplementary Fig. 14d).

We then compared autonomous robotic vascular access to manual cannulation without image guidance, with NIR image guidance and with US image guidance. We employed a full factorial experiment to evaluate the response to variation in the four tissue parameters identified previously (Supplementary Fig. 15). Mean first-attempt success rates for manual unassisted, manual NIR-guided and manual US-guided vascular access across these challenging conditions were 52.7%, 59.3% and 68.4%, respectively (Fig. 5c). Robotic vascular access resulted in 88.2% first-attempt success—a significant increase from the manual success rates (Kruskal–Wallis one-way ANOVA with Dunn's post hoc test,  $P<0.0001$  for all pairwise comparisons between robotic and control groups). Similarly, the robotic approach reduced the mean number of failed cannulation attempts per procedure ( $P<0.01$  for all pairwise comparisons to control groups) and the total time to access ( $P<0.0001$  for all pairwise comparisons to control groups). Compared to manual cannulation, the robotic approach demonstrated greater consistency in performance across the spectrum of tissue conditions, with the largest gains seen in the most difficult conditions (Fig. 5d and Supplementary Figs. 16–19). Robotic cannulation also reduced, among the total failed vascular access attempts, the percentage of unintended punctures of the posterior wall, as identified by visualization under B-mode US (Fig. 5c,d). Finally, retrospective analysis of US frames acquired during robotic cannulation demonstrated the possibility of automatically distinguishing successful and failed vascular access attempts based on the observed displacement of the vessel and predicted location of the needle tip at the time of puncture (Fig. 5e,f and Supplementary Fig. 20).



**Fig. 3 | Comparison of deep learning and expert assessment of upper-extremity forearm vessels.** **a**, Comparison of FCN and Rec-FCN performance for NIR stereo arm surface reconstruction, NIR vessel segmentation, and duplex US vessel segmentation from images acquired on left and right forearms of 22 volunteers. **b**, Vessel segmentation and disparity estimation from NIR image sequences based on Rec-FCN predictions. Dice, Sørensen-Dice coefficient; MHD, modified Hausdorff distance (for definitions see Supplementary Method 1). **c**, Vein and artery segmentation from B-mode and CDI sequences based on Rec-FCN predictions. **d**, Differentiation of arteries and veins from B-mode and CDI sequences. Also reported are the resulting classification thresholds based on maximization of F1 scores. Background pixels are excluded from analysis. **e,f**, Receiver operating characteristic (ROC) (**e**) and precision-recall (**f**) curves comparing the reliability of different methods of vessel classification (for details see Supplementary Fig. 10). Triangles indicate positions of maximum predictive performance based on Youden's J statistic (left) and F1 score (for definitions see Supplementary Method 2). PPV, positive predictive value; TPR, true positive rate; FPR, false positive rate. \*Kruskal-Wallis one-way ANOVA with Dunn's post hoc test.



**Fig. 4 | Real-time robotic tracking and motion compensation.** **a**, Time-series images of robotic vessel tracking on a volunteer. Insets show upper extremity vessels segmented from NIR image acquired prior to robotic tracking. Blue circles, tracked cannulation targets for different vessel branches; white arrows, desired robotic needle pose at each target. Adapted from ref. <sup>38</sup>. **b**, Representative trajectories observed during the first 60s (left) and first 5s (right) of NIR- and US-guided robotic tracking in comparison to ground-truth target trajectories. **c**, Absolute translational ( $|\varepsilon_{\text{trans}}|$ ) and rotational ( $|\varepsilon_{\text{rot}}|$ ) errors observed during robotic tracking, shown in comparison to ground-truth motions ( $|\nu_{\text{trans}}|$  and  $|\nu_{\text{rot}}|$ ). Errors are defined by magnitude differences between robotic and target positions in each frame. Ground-truth motion is defined as the magnitude velocity of the cannulation target in each frame, determined from manual annotation.  $|\hat{\varepsilon}|$  and  $|\hat{\nu}|$  indicate upper quartiles of error and ground-truth distributions, respectively. **d**, Correlation of predicted frame-to-frame translations ( $|\hat{\nu}_{\text{trans}}|$ ) and rotations ( $|\hat{\nu}_{\text{rot}}|$ ) to ground-truth motions ( $|\nu_{\text{trans}}|$  and  $|\nu_{\text{rot}}|$ ). Also shown are upper quartiles of predicted ( $|\hat{\nu}_{\text{NIR}}^{q3}|$  and  $|\hat{\nu}_{\text{US}}^{q3}|$ ) and true ( $|\nu_{\text{NIR}}^{q3}|$  and  $|\nu_{\text{US}}^{q3}|$ ) velocity distributions under NIR and US guidance.  $R^2$ , coefficient of determination. **e**, ROC (left) and precision-recall (right) curves indicating the reliability of classifying sudden arm movements based on predicted frame motions. Individual curves reflect varying velocity cutoff values defining sudden movement. In the two highlighted curves, velocity cutoffs were set to the upper quartiles of the translational and rotational velocity distributions ( $|\hat{\nu}_{\text{NIR}}^{q3}|$  and  $|\hat{\nu}_{\text{US}}^{q3}|$ ). Triangles indicate positions of maximum predictive performance based on Youden's J statistic (left) and F1 score (right). TPR, true positive rate; FPR, false positive rate. **f**, Reliability of detecting sudden arm movements (exceeding the upper quartiles of observed velocity distributions) during NIR (left) and US (right) guidance, based on four binary classification metrics (for definitions see Supplementary Method 2).

**In vivo autonomous blood drawing and fluid delivery in submillimetre vessels in rat models.** To evaluate in vivo vascular access in submillimetre vessels, we conducted image-guided robotic cannulations on lateral tail veins of 20 fully anaesthetized adult rats (13 white-coated (WC) and seven black-coated (BC)) and compared the performance to manual cannulation. Rat tail vein cannulations were chosen because the diameter ( $0.75 \pm 0.21$  mm) and depth ( $1.35 \pm 0.52$  mm) of the vessels were comparable to measurements in paediatric populations<sup>14</sup>. NIR and US imaging (Fig. 6a,b and Supplementary Fig. 21) increased the percentage and path length of detected vessels compared to manual assessment under visible light (Fig. 6c,d).

To compare manual and autonomous cannulation in the animals, 94 cannulation trials (24 unassisted manual, 23 NIR-guided manual, 21 US-guided manual and 31 robotic) were carried out on the 20 anaesthetized rats along with collection of 250 µl of blood and delivery of 250 µl saline bolus (Fig. 6e,f). Autonomous robotic cannulation improved the rates of first-attempt access (87.1% for the device versus 58.3% for unassisted manual (Kruskal–Wallis one-way ANOVA with Dunn's post hoc test,  $P=0.016$ ), 69.6% for NIR-guided manual ( $P=0.097$ ) and 61.9% for US-guided manual ( $P=0.036$ ; Fig. 6g). Similar improvements were observed in the success rates of blood collection and fluid delivery. Finally, robotic cannulation reduced the number of failed cannulation attempts (Fig. 6h), mean completion time per trial (Fig. 6i) and percentage of detected posterior wall punctures (Fig. 6j) compared to manual techniques.

## Discussion

Over 90% of diagnostic and therapeutic procedures in the emergency room, intensive care unit, catheterization lab and operating room require gaining vascular access<sup>8,13</sup>. Approximately one billion vascular access procedures are performed annually in the United States (and approximately four billion procedures worldwide)<sup>13</sup>. We focused on cannulation of upper-extremity peripheral vessels, which are the most common targets for cannulation (>95% of total procedures<sup>8</sup>) and considered particularly challenging due to their small size (typically 2 to 3 mm in adults and <1 mm in children) and tendency to roll or collapse<sup>44</sup>. In our studies, autonomous robotic cannulation reduced the mean number of failed access attempts by sixfold (1.8 per trial to 0.3 per trial) and increased first-stick success rates from 53% to 88% compared to blind manual access, with the largest gains seen in the most difficult physiological conditions (Fig. 5c,d). Moreover, we found that variance in robotic performance in the presence of physiological variability was lower than the variance of manual cannulation with or without image guidance. We posit that, by lowering the likelihood of failed attempts, robotic cannulation

could prevent injuries from multiple sticks, reduce complication rates and minimize the need for central catheter placement following unsuccessful peripheral access<sup>15,45</sup>. The present work provides motivation for further assessment of clinical benefits and risks across a broader demographic spectrum that includes both normal and difficult populations.

Safety remains a central concern in the development of medical robots, particularly those with potential autonomous capabilities. We described several mechanisms to reduce the risks of injury during vascular access. We showed that sudden arm motions could be detected from the NIR and US sequences (Fig. 4d–f), and have coupled the needle to a 5 N load sensor that continuously measures axial forces at the needle tip (Supplementary Fig. 2b–e). We designed the device to electromagnetically release the needle when sudden motions or excessive insertion forces are detected. Furthermore, we observed in vitro that failed access attempts (including posterior wall punctures, which are a common cause of access-related injuries such as extravasation and arteriovenous fistula creation<sup>12</sup>) could be recognized based on analysis of vessel motion and needle tip position in the US image at the time of cannulation (Fig. 5e,f and Supplementary Fig. 20). Future studies will extend these characterizations in vivo and evaluate the reliability of these combined safety mechanisms in the clinical setting.

In the current device, robot trajectories are updated according to the vessel pose computed from deep neural network predictions. However, rather than using deep learning to control robot motions in an end-to-end manner, trajectories are determined from the robot kinematic parameters based on set control policies. It is possible, however, that such policies should be context-specific. Learning-based strategies, for instance based on recurrent networks or reinforcement paradigms<sup>46</sup>, have shown promise in producing optimal policies for robotic surgical tasks<sup>47</sup>. Subsequent efforts will investigate whether learned trajectories result in better performance outcomes compared to traditional methods of robotic control.

Full end-to-end autonomy will require the ability to recognize suitable vascular access locations, define secondary sites when initial attempts fail, select proper needle and catheter sizes, and adapt to individual differences in anatomy and physiological status. In our NIR imaging studies, we demonstrated that, in a majority of cases, it was possible to identify vessel branches deemed appropriate for cannulation based on predictive analysis of combined image features extracted from deep learning-based segmentations (Supplementary Fig. 8). Similar observations have been reported in previous studies<sup>20,36,48</sup>. Nevertheless, human experts rely also on anatomical and clinical knowledge when defining a strategy for vascular access and make judgements based on information not easily described by low-level descriptors alone. Future efforts will investigate methods

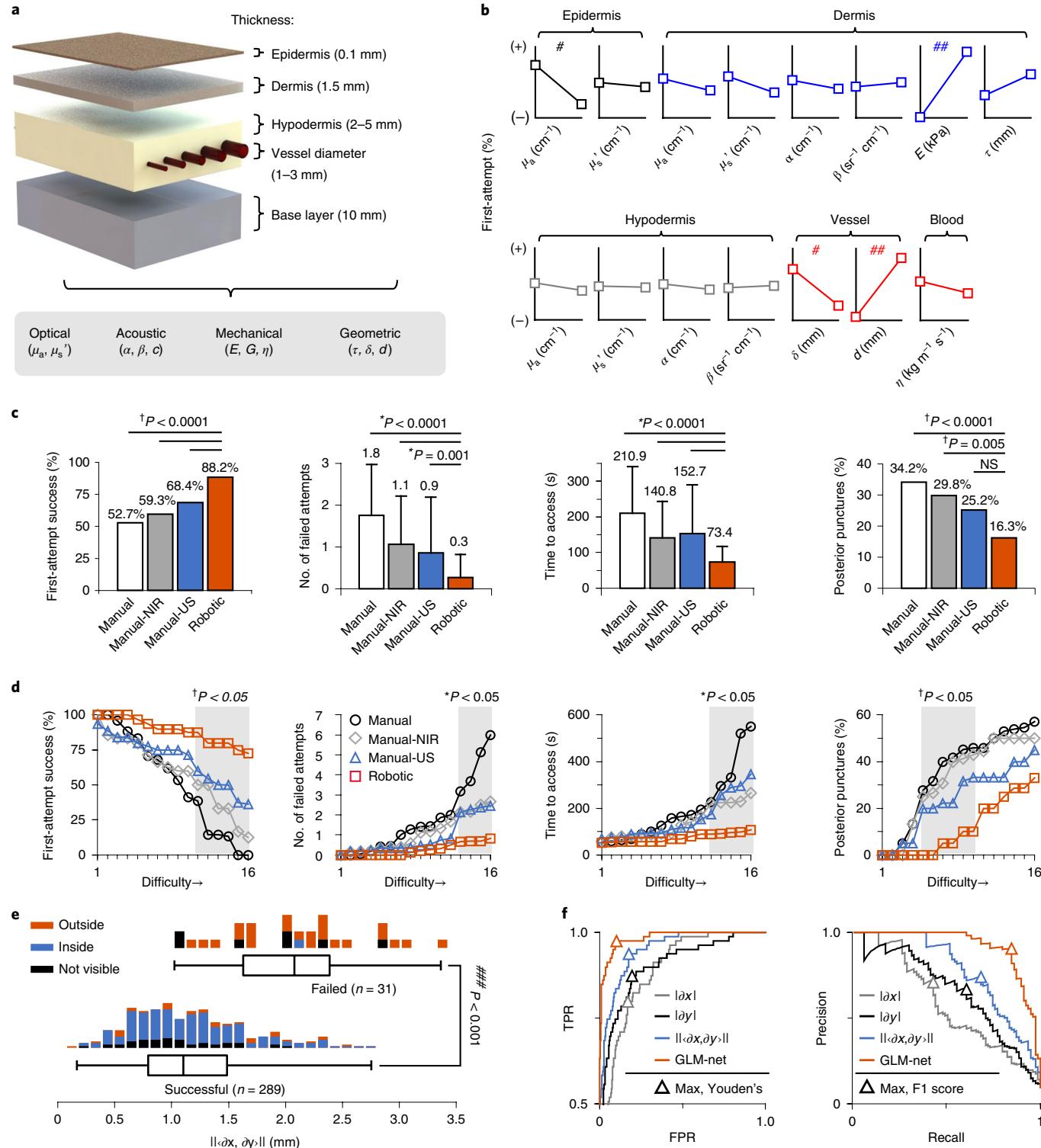
**Fig. 5 | Autonomous cannulation in tissue-mimicking models over a broad demographic spectrum.** **a**, Multilayered models simulating optical (absorption  $\mu_a$ , scattering  $\mu_s'$ ), acoustic (attenuation  $\alpha$ , backscattering  $\beta$ , sound speed  $c$ ), mechanical (elasticity  $E$ , shear modulus  $G$ , viscosity  $\eta$ ) and geometric (tissue thickness  $\tau$ , vessel depth  $\delta$ , vessel diameter  $d$ ) properties of human skin and vessels<sup>43</sup>. **b**, The main effects of two-level fractional factorial evaluation of robotic first-attempt success across 15 tissue conditions (Supplementary Fig. 14c,d). **c**, Comparison of unassisted manual ( $n=160$  total trials), NIR-guided manual ( $n=160$ ), US-guided manual ( $n=160$ ) and robotic ( $n=320$ ) cannulation in a full factorial (four-variable, two-level) experiment consisting of 16 independent combinations of tissue conditions (Supplementary Fig. 15). Posterior wall puncture rate is reported as the percent of failed attempts per trial. Subgroup analyses are reported in Supplementary Figs. 16–19. **d**, Cannulation performance across the 16 tissue conditions, shown in order of increasing difficulty. Grey shaded regions indicate conditions for which robotic cannulations significantly outperformed manual trials. **e**, Vessel magnitude displacements  $||\langle dx, dy \rangle||$  (mm) and estimated final needle tip locations,  $n_{\text{final}} \in \{n_{\text{inv}}, n_{\text{out}}, n_{\text{none}}\}$  ( $n_{\text{inv}}$ , needle tip inside vessel;  $n_{\text{out}}$ , needle tip outside vessel;  $n_{\text{none}}$ , needle tip not detected) in successful and failed robotic cannulation attempts across 16 tissue conditions. Needle tip location is estimated based on image analysis using a statistical model-fitting algorithm (Supplementary Fig. 20a). Successful access determined by observation of blood flash in the needle hub following cannulation. **f**, ROC (left) and precision-recall (right) indicating reliability of predicting failed punctures. Predictors: absolute lateral vessel displacement  $|\delta x|$  (mm), absolute axial vessel displacement  $|\delta y|$  (mm), vessel magnitude displacement  $||\langle dx, dy \rangle||$  (mm), and predictions from an elastic-net regularization generalized linear model GLM-net ( $\delta x, \delta y, n_{\text{final}}$ ) (for details see Supplementary Fig. 20b). Triangles indicate positions of maximum predictive performance based on Youden's J statistic (left) and F1 score (right) (see definitions in Supplementary Method 2). TPR, true positive rate; FPR, false positive rate.  $^{\#}P < 0.05$ ,  $^{**}P < 0.01$ ,  $^{***}P < 0.001$ , unpaired t-tests with Bonferroni correction (**b,e**). \*One-way ANOVA with Tukey's honestly significant difference (HSD) test (**c,d**; bars show mean  $\pm$  s.d.). <sup>†</sup>Kruskal–Wallis one-way ANOVA with Dunn's post hoc test (**c,d**; bars show percentages).

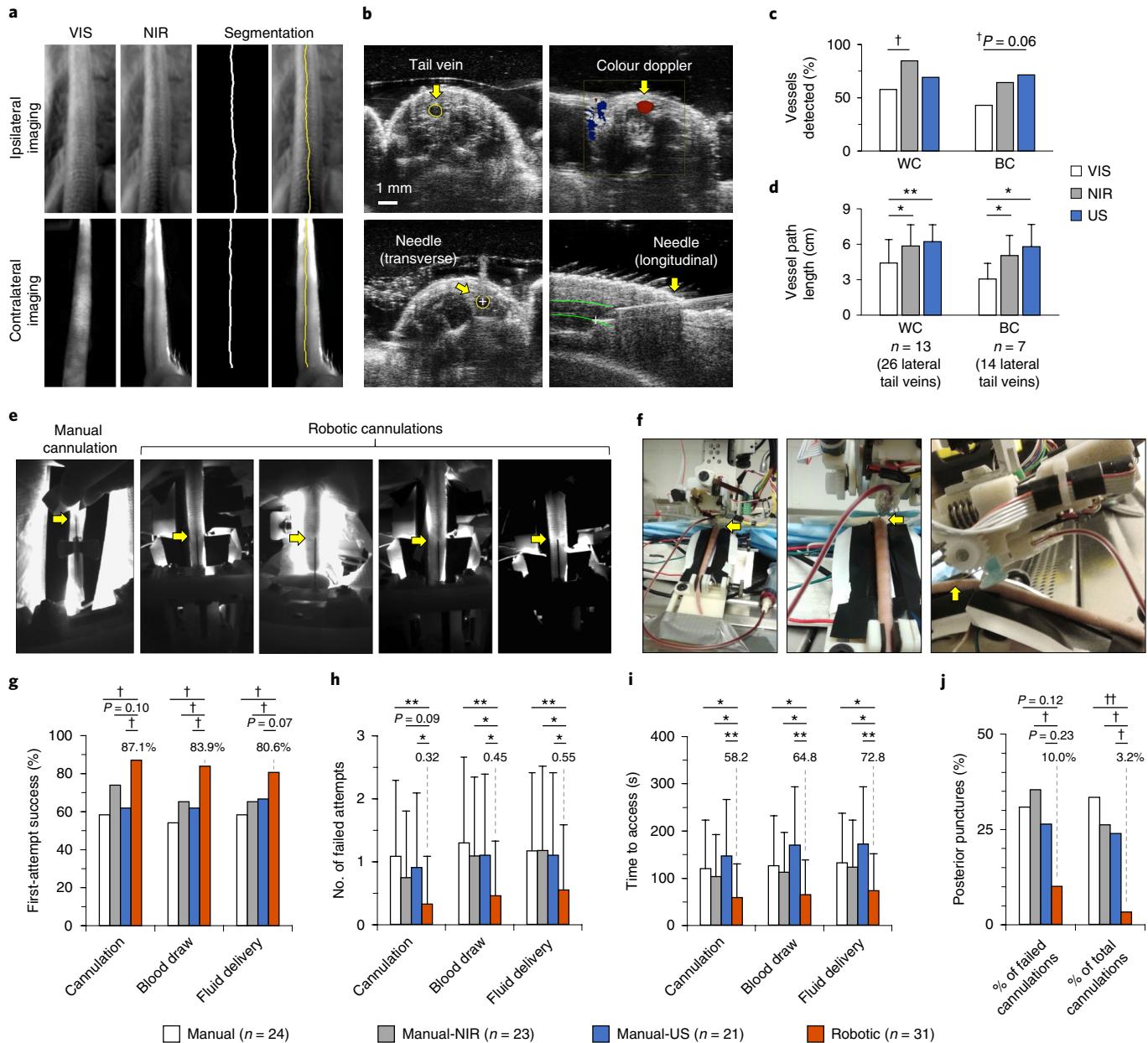
to capture richer anatomical representations and test whether an autonomous system utilizing such methods can determine appropriate cannulation routes reliably.

The robotic paradigm may be extended to address clinical challenges in minimally invasive endovascular workflows, where accurate cannulation of major vessels (such as the common femoral artery) is a prerequisite to surgical success and where repeat punctures increase the risk of arterial trauma and haemorrhage<sup>15,25</sup>. Outside the hospital, robotic technologies could allow emergency medical providers to obtain rapid vascular access under time-critical

conditions and bring advanced interventional and resuscitation capabilities to remote and resource-limited environments<sup>11</sup>. Finally, the device has the potential to serve as a platform to merge automated phlebotomy and diagnostic blood analysis, facilitating the provision of critical haematological information at the point of the blood draw<sup>49,50</sup>.

In summary, the present study demonstrates the preclinical feasibility of autonomous, image-guided robotic vascular access, blood drawing and fluid delivery. The findings provide evidence that, by exploiting the capacity of modern deep networks to encode





**Fig. 6 | In vivo autonomous blood drawing and fluid delivery in submillimetre vessels of rats.** **a**, Rat lateral tail vein imaging under visible (VIS, 453 nm) and near-infrared (NIR, 914 nm) light. Top images: light source ipsilateral to detector, image formation from backscattered light. Lower images: light source contralateral to detector, image formation from ballistic light. **b**, Robotic tail vein cannulation under US guidance. **c,d**, Percentage (**c**) and path length (**d**) of lateral tail veins detected by VIS, NIR and US vein imaging in 13 WC and 7 BC rats (26 and 14 lateral tail veins, respectively). **e**, Real-time visualization of lateral tail veins under NIR imaging during manual and robotic cannula insertion. **f**, Automatic tail vein cannulation and blood collection using the robotic device. **g–j**, Comparison of unassisted manual, NIR-guided manual, US-guided manual, and robotic cannulation, blood drawing and fluid delivery in tail veins of anaesthetized rats. \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ , one-way ANOVA with Tukey's HSD test (**d,h,i**; bars show mean  $\pm$  s.d.). † $P < 0.05$ , ‡ $P < 0.01$ , †† $P < 0.001$ , Kruskal–Wallis one-way ANOVA with Dunn's post hoc test (**g,j**; bars show percentages). Yellow arrows indicate cannulation targets (**b,e,f**).

multimodal spatiotemporal imaging information, autonomous systems may be able to outperform human experts on challenging visuomotor tasks in dynamic environments. Such systems, if successfully translated, offer the possibility of reducing injuries, improving procedural efficiency and outcomes, carrying out tasks with minimal supervision when resources are limited, and allowing human attention to be dedicated to other critical aspects of medical care.

## Methods

**Deep learning architecture.** The deep Rec-FCN models trained for NIR and US guidance are based on a neural network architecture that embeds a recurrent

block within a U-net-like<sup>40,41,51</sup> fully convolutional encoder–decoder network (Fig. 2c,d and Supplementary Figs. 3–5). We applied stride-2 convolutions using  $3 \times 3$  kernels for downsampling in the encoder and corresponding transpose convolutions for upsampling in the decoder. Each convolutional layer is preceded by batch normalization<sup>52</sup> and followed by nonlinear activation using parametric rectified linear units<sup>53</sup>. Residual connections (identity mappings based on element-wise summation)<sup>54</sup> are introduced within each convolutional block to lessen the potential influence of vanishing gradients. We incorporated skip connections (concatenation along channel dimensions)<sup>40</sup> between the encoder and decoder layers to combine semantic information across feature resolutions. Similarly, within the recurrent block, latent features produced by the encoder are concatenated with features from previous time steps before being passed to the decoder. We evaluated three recurrent units (a standard recurrent neural

network (RNN)<sup>55</sup>, a convolutional long short-term memory (LSTM) unit<sup>56</sup> and a convolutional gated recurrent unit (GRU)<sup>57,58</sup> for temporal inference (Supplementary Fig. 5). We used convolutional GRU in the present work. We also evaluated a variant of the Rec-FCN architecture that incorporates a recurrent unit at each spatial resolution level within the encoder–decoder structure, but we did not observe significant differences in performance using this approach compared to models with recurrent connection at the innermost layer only.

The first network (Fig. 2c) was trained on previously acquired NIR video data of left and right forearm vessels from nine healthy participants<sup>59</sup>. Inputs into the network were two-channel tensors with dimensions  $384 \times 288 \times 2$ , with channels comprising rectified left and right stereo NIR images. The labels and outputs were three-channel tensors with dimensions  $384 \times 288 \times 3$ , with channels comprising the left segmentation, right segmentation and disparity maps, respectively. We used a multi-task loss function<sup>60–62</sup>

$$\text{Loss}_{\text{Overall}} = w_1 \text{Loss}_{\text{Generalized Dice}}^{\text{Segmentation}} + w_2 \text{Loss}_{\text{Weighted Cross Entropy}}^{\text{Segmentation}} \\ + w_3 \text{Loss}_{\text{Mean Squared Error}}^{\text{Disparity}} + w_4 \text{Loss}_{\text{Total Variation}}^{\text{Disparity}}$$

with weights  $w_1 = 0.5$ ,  $w_2 = 0.1$ ,  $w_3 = 0.3$  and  $w_4 = 0.1$  such that the segmentation loss terms were applied only to the segmentation output channels and the disparity loss terms were applied only to the disparity output channel. For network training, NIR video acquisitions were split into 0.5 s sequences, each containing 15 frames, and introduced into the deep learning model alongside manually generated segmentation and disparity labels (for annotation details see Supplementary Method 4). During testing, the model operated on continuous sequences. In total, 186 short sequences (2,790 annotated frames) were used in training and validation, and 22 full sequences (8,912 acquired frames, of which 595 were annotated) were used in testing.

The second network (Fig. 2d) was pre-trained on two public US image datasets<sup>63,64</sup> and fine-tuned on previously acquired transverse 2D US images of peripheral upper-extremity vessels from nine subjects using two clinical US transducers (L18-10L30H-4, Telemed Ultrasound and SeeMore Near-Field 7.5/24 MHz, Interson)<sup>50</sup>. Inputs were two-channel tensors with dimensions  $512 \times 416 \times 2$  containing B-mode and CDIs. In the color Doppler channel, we applied a five-fold increase in image gain to venous (negative) flow velocities. The loss function was given by<sup>60,61</sup>

$$\text{Loss}_{\text{Overall}} = w_1 \text{Loss}_{\text{Generalized Dice}}^{\text{Segmentation}} + w_2 \text{Loss}_{\text{Weighted Cross Entropy}}^{\text{Segmentation}}$$

with weights  $w_1 = 0.8$  and  $w_2 = 0.2$ . The loss terms were applied to the output channels individually and then summed across the channels. As before, videos were split into 0.5 s sequences containing 11 frames each, which resulted in the inclusion of 278 short sequences (4,326 annotated frames) along with manual ground-truth segmentation labels for training and validation (see annotation details in Supplementary Method 4). In testing, 22 full sequences (15,342 acquired frames, including 1,018 annotated frames) were used.

For both networks, we applied standard data augmentation techniques including random rotation, horizontal and vertical flips, cropping and scaling, and gain and contrast adjustment. We also performed temporal augmentation by alternating the direction of the sequences between epochs and by applying window warping (randomly adjusting the time step between frames within each sequence to effectively speed up or slow down the motion)<sup>65</sup>. Each network was trained end-to-end using the standard stochastic gradient descent algorithm with the Adam optimizer<sup>66</sup> and L2 weight decay regularization<sup>57</sup>. The networks were implemented using the Tensorflow library<sup>68</sup>.

**Bimodal NIR and US vessel imaging.** In the NIR guidance step, the robot infers from the first trained network to segment and compute the 3D pose of peripheral vessels. Arm motions are estimated based on frame-to-frame non-rigid registration of the predicted segmentations. The robot positions the US transducer against the arm surface based on the predictions of the first network while adjusting to the estimated arm motions. Because the depth of the arm surface is computed from the NIR stereo image sequences, the device is able to position the transducer against the skin and maintain acoustic contact without applying excess pressure, which could lead to vessel compression. In the US guidance step, the second network segments the vessel lumen and classifies each segmented vessel as either a vein or artery based on B-mode and colour Doppler sequences. As before, frame-to-frame motion is estimated from the sequence of segmentation predictions and used to continuously update the robot trajectory before and during cannulation.

The NIR light source comprises an array of 15 LEDs each with dimensions of  $2.0 \times 1.5$  mm. We used LEDs with wavelengths centred at 757 and 914 nm to maximize the optical absorption in blood while minimizing absorption from water and fat<sup>21</sup>. Two miniature CMOS cameras (VRMUsb12, VRMagic) were geometrically calibrated to form a stereo vision system that acquires pairs of  $752 \times 480$  images at 31 Hz. In the stereo image rectification step, the images are downsampled to  $384 \times 288$ . Each camera is coupled with a wide-angle ( $120^\circ$ ) lens and a polarizing filter. The filter is oriented orthogonal to a second set of polarizers above the LED arrays to eliminate specular reflections at the skin surface based on cross-polarization gating.

In the phantom and human studies, we used an 18 MHz clinical linear-array US transducer (L18-10L30H-4, Telemed) with a longitudinal imaging field of 3 cm, maximum depth of 5 cm and focal depth of 1 cm. B-mode images and CDIs were acquired at 22 Hz. We extended the Doppler region of interest to approximately the width of the B-mode image size and fixed the Doppler angle and gain. In the in vivo studies, we compared imaging quality between the 18 MHz clinical transducer (Supplementary Fig. 21a) and a high-frequency 32 MHz transducer (Veo MS550S, VisualSonics; Supplementary Fig. 21b). The higher frequency resulted in improved visualization of vessels with diameters as low as 300  $\mu\text{m}$  and at depths of up to 15 mm. The increased frequency also improved the sensitivity of CDI in submillimetre vessels.

**Robotic system design.** The device (Fig. 1 and Supplementary Fig. 1) is  $1.0 \times 0.9 \times 0.7$  cubic feet in volume and weighs 3 kg. The robotic system consists of a 6-DOF base positioning system and a 3-DOF distal manipulator, resulting in 9 DOF in total. The base positioning system serves to position and orient the distal manipulator along the skin surface, orient the cameras and align the US transducer and needle with the target vessel. The 3-DOF manipulator (0.5 kg weight,  $7 \times 3 \times 8$  cm volume) couples the NIR imaging system, US transducer and motorized insertion mechanism into a single modular unit and is mounted directly to the base positioning system. Forces at the needle tip are measured with a 5 N uniaxial force sensor (FSG-5, Honeywell) integrated in the 3-DOF manipulator (Supplementary Fig. 2b). A routine was implemented for simultaneous calibration of the intrinsic camera and US image parameters, extrinsic camera-to-robot and US-to-robot parameters, and robot joint parameters<sup>69</sup>. Real-time software for image-guidance and robotic control were developed in LabVIEW and C++ and utilized open-source libraries including OpenCV<sup>70</sup>, Point Cloud Library<sup>71</sup> and ITK<sup>72</sup>. Further details of the system design are described in Supplementary Method 3.

**Vessel imaging and robotic tracking on healthy volunteers.** Twenty-two healthy volunteers, 18 years of age and older, were recruited following approval by the Rutgers University Institutional Review Board. The left and right forearm, wrist and hand of each volunteer were imaged by the device under NIR illumination. Retrospective analysis identified six upper-extremity superficial venous branches consistently across participants (Fig. 2e). Ground-truth segmentations were generated on a frame-by-frame basis. For each subject, one resulting ground-truth segmentation frame from each detected vessel branch was confirmed by expert review. Arterial branches were not evaluated, as the increased depth of the arteries (typically  $>5$  mm beneath the skin surface) limited their visibility under NIR light.

Transverse 2D US imaging identified six upper-extremity venous branches and three upper-extremity arterial branches across participants (Fig. 2f). Ground-truth segmentations were generated manually, with one resulting segmentation frame from each detected vessel confirmed by expert review. The reviewed frames were then used to evaluate binary classification of veins and arteries (Fig. 3d,e). For colour Doppler acquisitions, clutter signals were suppressed using finite impulse response filtering with default upper-extremity venous presets provided by the commercial US systems<sup>42</sup>. Methods for ground-truth annotation are described in Supplementary Method 4.

Thirteen volunteers were subsequently enrolled in the robotic tracking study. To measure tracking errors over varying speeds, each volunteer was asked to move his or her arm inside the device workspace for 60 s with random motions (Fig. 4a). A target cannulation site, selected randomly from among the detected vessels, was robotically tracked under NIR and US image guidance. Frame-to-frame motion along the segmented vessel centrelines was estimated based on deformable point set registration<sup>73</sup> (Supplementary Fig. 7). In the NIR-guided tracking phase, the objective was to maintain the target vessel at the centre of the NIR imaging field. Similarly, the robot looked to centre the target vessel within the US image during US-guided tracking. The ground-truth positions of the vessel target across all NIR and US image frames were retrospectively determined with manual confirmation. Ground-truth vessel positions were then compared to the robotic tracking positions to compute tracking errors (Fig. 4b,c and Supplementary Fig. 12). We did not evaluate robotic needle insertion accuracy in these studies, as the robot did not cannulate through the skin.

**In vitro studies on tissue-mimicking models.** The randomized in vitro studies consisted of three sets of experiments. First, we optimized three robotic cannulation parameters (insertion angle ( $15^\circ$  and  $30^\circ$ ), insertion speed (1 and  $10 \text{ mm s}^{-1}$ ) and needle size (21 and 25 G)) using tissue-mimicking models (Supplementary Fig. 14a). Needle lengths of 1 inch were used in all phantom cannulation trials. The results of the optimization (Supplementary Fig. 14b) were used in defining cannulation settings for the subsequent in vitro and in vivo experiments.

Second, we investigated robotic cannulation across 15 physiological and anatomical parameters simulated using in vitro tissue-mimicking models (Fig. 5a,b). We employed a fractional factorial experimental design (specifically, an L16 ( $2^{15}$ ) orthogonal array of parameter combinations; Supplementary Fig. 14c,d) that constrains the input parameters to be orthogonal to one another and ensures an unbiased, uniform and maximally efficient sampling of the parameter space<sup>74</sup>. Ten replicate cannulation trials were carried out per experimental condition in

randomized order, along with withdrawal of 1 ml of blood-mimicking solution and intravenous delivery of 1 ml saline. The physiological ranges of each tissue parameter are summarized in Supplementary Fig. 13 and described previously<sup>43</sup>.

Third, we compared autonomous robotic cannulation ( $n=320$  total trials) to manual cannulation without image guidance ( $n=160$  trials), with NIR guidance ( $n=160$  trials) and with US guidance ( $n=160$  trials) (Fig. 5c,d and Supplementary Figs. 15–19). We employed a full factorial experiment using 16 different models (Supplementary Fig. 15a). We evaluated the four tissue parameters observed to most strongly influence device performance in the earlier fractional factorial studies (Fig. 5b). Specifically, the models encompassed two epidermis absorption coefficients (5 and  $45\text{ cm}^{-1}$ , measured at  $914\text{ nm}$ ), two hypodermis elasticities (5 and  $25\text{ kPa}$ ), two hypodermis thicknesses (2 and 5 mm) and two vessel diameters (1 and 3 mm). Based on the results of the first set of optimization experiments (Supplementary Fig. 14b), we used 25G needles of 1 inch length, device insertion angle of  $30^\circ$  and insertion speed of  $10\text{ mm s}^{-1}$ . For manual cannulations, 10 randomized, replicate trials were performed per tissue condition, with cannulation, blood draw and saline delivery as endpoints. For robotic cannulations, 20 trials were performed per condition. The manual cannulations were performed by a senior research fellow who had undergone standard and US-guided vascular access training prior to the study. An unbiased orthogonal subset of manual trials were then repeated by a clinical expert, with retrospective statistical analysis indicating equivalence in cannulation performance between the research and clinical operators (Supplementary Fig. 22). Operators were allowed five practice attempts on each tissue-mimicking model before data collection.

**In vivo studies on rats.** Thirteen WC male Sprague–Dawley rats and seven BC male Sentinel rats (weight  $260.9 \pm 75.3\text{ g}$ ) were included in the in vivo study following approval by the Rutgers University Institutional Animal Care and Use Committee. The study was carried out over a four-month period with no more than eight repeated trials per animal. Each rat was anaesthetized by 5% isoflurane gas administered by inhalation and subsequently maintained under anaesthesia for up to 1 h by 2.5% isoflurane gas. Once anaesthetized, each animal was positioned on a raised platform mounted to the device that secures the tail. A water recirculating blanket was placed underneath to maintain blood flow reduction and prevent hypothermia. Before the procedure, the tail was cleaned with 70% ethanol and soaked in  $40^\circ\text{C}$  water for 1 min to induce vasodilation. A tourniquet was applied 1 cm from the proximal end of the tail.

Two methods of NIR tail vessel imaging were compared (Fig. 6a). In the first method, the NIR LED light source was arranged ipsilateral to the CMOS cameras to provide reflectance-based illumination of the tail. In the second method, the light source was positioned contralateral to the cameras to allow transmittance imaging. Finally, US image qualities at two frequencies (18 and 32 MHz) were compared, with the higher-frequency imaging observed qualitatively to result in improved resolution of submillimetre vessels (Fig. 6b and Supplementary Fig. 21).

The randomized in vivo cannulations were carried out using transmittance NIR imaging at 757 and  $914\text{ nm}$  followed by longitudinal and transverse duplex US imaging at 32 MHz. Cannulation sites were manually determined, and a spacing of at least 3 cm between cannulations was maintained along the vessel, starting from the distal end of the tail. For the robotic trials, the position of the segmented vessel closest in distance to the defined cannulation site was then used as the target for access. Cannulation success was defined by the visual observation of the needle tip within the vessel lumen on the US image and the presence of blood flash in the hub of the infusion cannula. In cases where blood flash was not observed, lumen access was confirmed by threading a sterile 0.1 mm-diameter wire filament to the tip of the needle and visualizing the distal end of the wire within the lumen. Successful cannulation was followed by collection of  $250\text{ }\mu\text{l}$  of blood and infusion of  $250\text{ }\mu\text{l}$  saline. We used 27G needles (0.75 inch length) in all trials to allow cannulation of vessels approaching 0.5 mm diameter and to minimize potential vessel damage. All robotic cannulations were performed with an insertion angle of up to  $30^\circ$  and insertion speed of  $10\text{ mm s}^{-1}$ . All manual cannulations were performed by a senior research fellow with three months prior training in rodent and small animal venipuncture.

**Reporting Summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

Test datasets for evaluating source code are available at <https://github.com/alvchn/nmi-vasc-robot>. Public data used in the study are available in the SPLab Ultrasound Image Database ([http://splab.cz/wp-content/uploads/2014/05/ARTERY\\_TRANSVERSAL.zip](http://splab.cz/wp-content/uploads/2014/05/ARTERY_TRANSVERSAL.zip) and [http://splab.cz/wp-content/uploads/2013/11/us\\_images.zip](http://splab.cz/wp-content/uploads/2013/11/us_images.zip)), the PICMUS Database ([https://www.creatis.insa-lyon.fr/Challenge/IEEE\\_IUS\\_2016/download](https://www.creatis.insa-lyon.fr/Challenge/IEEE_IUS_2016/download)) and the SPLab Tecnocampus Hand Image Database (<http://splab.cz/en/download/databaze/tecnocampus-hand-image-database>).

## Code availability

Source code are available from the Github repository: <https://github.com/alvchn/nmi-vasc-robot>. Use of the code is subject to a limited right to use for academic, governmental or not-for-profit research. Use of the code for commercial or clinical purposes is prohibited in the absence of a Commercial License Agreement from

Rutgers, The State University of New Jersey. References to open-source software used in the study are provided within the paper.

Received: 1 July 2019; Accepted: 9 January 2020;

## References

- Yang, G. Z. et al. Medical robotics—regulatory, ethical, and legal considerations for increasing levels of autonomy. *Sci. Robot.* **2**, eaam8638 (2017).
- Moustris, G. P., Hiridis, S. C., Deliparaschos, K. M. & Konstantinidis, K. M. Evolution of autonomous and semi-autonomous robotic surgical systems: a review of the literature. *Int. J. Med. Robot. Comput. Assist. Surg.* **7**, 375–392 (2011).
- Shademan, A. et al. Supervised autonomous robotic soft tissue surgery. *Sci. Transl. Med.* **4**, 337 (2016).
- Edwards, T. L. et al. First-in-human study of the safety and viability of intracarotid robotic surgery. *Nat. Biomed. Eng.* **2**, 649–656 (2018).
- Fagogenis, G. et al. Autonomous robotic intracardiac catheter navigation using haptic vision. *Sci. Robot.* **4**, eaaw1977 (2019).
- Weber, S. et al. Instrument flight to the inner ear. *Sci. Robot.* **2**, eaal4916 (2017).
- Daudelin, J. et al. An integrated system for perception-driven autonomy with modular robots. *Sci. Robot.* **3**, eaat4983 (2018).
- Niska, R., Bhuiya, F. & Xu, J. National hospital ambulatory medical care survey: 2010 emergency department summary. *Natl Health Stat. Report* **2010**, 1–31 (2010).
- Horattas, M. C. et al. Changing concepts in long-term central venous access: catheter selection and cost savings. *Am. J. Infect. Control* **29**, 32–40 (2001).
- Sampalis, J. S., Lavoie, A., Williams, J. I., Mulder, D. S. & Kalina, M. Impact of on-site care, prehospital time, and level of in-hospital care on survival in severely injured patients. *J. Trauma* **32**, 252–261 (1993).
- Hulse, E. J. & Thomas, G. O. Vascular access on the 21st century military battlefield. *J. R. Army Med. Corps* **156**, 285–390 (2010).
- Armenteros-Yeguas, V. et al. Prevalence of difficult venous access and associated risk factors in highly complex hospitalised patients. *J. Clin. Nurs.* **26**, 4267–4275 (2017).
- Lamperti, M. & Pittiruti, M. II. Difficult peripheral veins: turn on the lights. *Br. J. Anaesth.* **110**, 888–891 (2013).
- Rauch, D. et al. Peripheral difficult venous access in children. *Clin. Pediatr. (Phila)* **48**, 895–901 (2009).
- Ortiz, D. et al. Access site complications after peripheral vascular interventions: incidence, predictors, and outcomes. *Circ. Cardiovasc. Interv.* **7**, 821–828 (2014).
- Lee, S. et al. A transparent bending-insensitive pressure sensor. *Nat. Nanotechnol.* **11**, 472–478 (2016).
- Chen, Z. et al. Non-invasive multimodal optical coherence and photoacoustic tomography for human skin imaging. *Sci. Rep.* **7**, 17975 (2017).
- Kolkman, R. G. M., Hondebrink, E., Steenbergen, W. & De Mul, F. F. M. In vivo photoacoustic imaging of blood vessels using an extreme-narrow aperture sensor. *IEEE J. Sel. Top. Quantum Electron.* **9**, 343–346 (2003).
- Matsumoto, Y. et al. Label-free photoacoustic imaging of human palmar vessels: a structural morphological analysis. *Sci. Rep.* **8**, 786 (2018).
- Meiburger, K. M. et al. Skeletonization algorithm-based blood vessel quantification using in vivo 3D photoacoustic imaging. *Phys. Med. Biol.* **61**, 7994–8009 (2016).
- Bashkatov, A. N., Genina, E. A., Kochubey, V. I. & Tuchin, V. V. Optical properties of human skin, subcutaneous and mucous tissues in the wavelength range from 400 to 2,000 nm. *J. Phys. D* **38**, 2543–2555 (2005).
- Paquit, V. C., Tobin, K. W., Price, J. R. & Mériadeau, F. 3D and multispectral imaging for subcutaneous veins detection. *Opt. Express* **17**, 11360–11365 (2009).
- Lamperti, M. et al. International evidence-based recommendations on ultrasound-guided vascular access. *Intensive Care Med.* **38**, 1105–1117 (2012).
- Egan, G. et al. Ultrasound guidance for difficult peripheral venous access: systematic review and meta-analysis. *Emerg. Med. J.* **30**, 521–526 (2013).
- Seto, A. H. et al. Real-time ultrasound guidance facilitates femoral arterial access and reduces vascular complications: FAUST (Femoral Arterial Access with Ultrasound Trial). *JACC Cardiovasc. Interv.* **3**, 751–758 (2010).
- Stolz, L. A., Stolz, U., Howe, C., Farrell, I. J. & Adhikari, S. Ultrasound-guided peripheral venous access: a meta-analysis and systematic review. *J. Vasc. Access* **16**, 321–326 (2015).
- Antoniou, G. A., Riga, C. V., Mayer, E. K., Cheshire, N. J. W. & Bicknell, C. D. Clinical applications of robotic technology in vascular and endovascular surgery. *J. Vasc. Surgery* **53**, 493–499 (2011).
- Zivanovic, A. & Davies, B. L. A robotic system for blood sampling. *IEEE Trans. Inf. Technol. Biomed.* **4**, 8–14 (2000).
- Cheng, Z. et al. A hand-held robotic device for peripheral intravenous catheterization. *Proc. Inst. Mech. Eng. H J. Eng. Med.* **231**, 1165–1177 (2017).

30. Kobayashi, Y. et al. Use of puncture force measurement to investigate the conditions of blood vessel needle insertion. *Med. Eng. Phys.* **35**, 684–689 (2013).
31. Kobayashi, Y. et al. Preliminary in vivo evaluation of a needle insertion manipulator for central venous catheterization. *Robomech. J.* **1**, 1–18 (2014).
32. Hong, J., Dohi, T., Hashizume, M., Konishi, K. & Hata, N. An ultrasound-driven needle-insertion robot for percutaneous cholecystostomy. *Phys. Med. Biol.* **49**, 441–455 (2004).
33. de Boer, T., Steinbuch, M., Neerken, S. & Kharin, A. Laboratory study on needle–tissue interaction: toward the development of an instrument for automated venipuncture. *J. Mech. Med. Biol.* **7**, 325–335 (2007).
34. Carvalho, P., Kesari, A., Weaver, S., Flaherty, P. & Fischer, G. Robotic assistive device for phlebotomy. In *Proc. ASME 2015 International Design and Engineering Technical Conferences & Computers and Information in Engineering Conference* Vol. 3, 47620 (ASME, 2015).
35. Brewer, R. *Improving Peripheral IV Catheterization Through Robotics—From Simple Assistive Devices to a Fully Autonomous System* (Stanford University, 2015).
36. Chen, A. I., Nikitczuk, K., Nikitczuk, J., Maguire, T. J. & Yarmush, M. L. Portable robot for autonomous venipuncture using 3D near infrared image guidance. *Technology* **1**, 72–87 (2013).
37. Harris, R., Mygatt, J. & Harris, S. System and methods for autonomous intravenous needle insertion. US patent 9,364,171 (2011).
38. Balter, M. L., Chen, A. I., Maguire, T. J. & Yarmush, M. L. Adaptive kinematic control of a robotic venipuncture device based on stereo vision, ultrasound, and force guidance. *IEEE Trans. Ind. Electron.* **64**, 1626–1635 (2017).
39. Lecun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
40. Long, J., Shelhamer, E. & Darrell, T. Fully convolutional networks for semantic segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**, 640–651 (2017).
41. Valipour, S., Siam, M., Jagersand, M. & Ray, N. Recurrent fully convolutional networks for video segmentation. In *Proc. 2017 IEEE Conference on Applications of Computer Vision* 26–36 (IEEE, 2017).
42. Bjaerum, S., Torp, H. & Kristoffersen, K. Clutter filter design for ultrasound color flow imaging. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **49**, 204–216 (2002).
43. Chen, A. I. et al. Multilayered tissue mimicking skin and vessel phantoms with tunable mechanical, optical, and acoustic properties. *Med. Phys.* **43**, 3117–3131 (2016).
44. Lewis, G. C., Crapo, S. A. & Williams, J. G. Critical skills and procedures in emergency medicine: vascular access skills and procedures. *Emerg. Med. Clin. North Am.* **31**, 59–86 (2013).
45. Galena, H. J. Complications occurring from diagnostic venipuncture. *J. Fam. Pract.* **34**, 582–584 (1992).
46. Mnih, V. et al. Human-level control through deep reinforcement learning. *Nature* **518**, 529–533 (2015).
47. Nguyen, N. D., Nguyen, T., Saeid, N., Bhatti, A. & Guest, G. Manipulating soft tissues by deep reinforcement learning for autonomous robotic surgery. In *Proc. 2019 IEEE International Systems Conference* 1–7 (IEEE, 2019).
48. Bullitt, E., Muller, K. E., Jung, I., Lin, W. & Aylward, S. Analyzing attributes of vessel populations. *Med. Image Anal.* **9**, 39–49 (2005).
49. Balter, M. L. et al. Automated end-to-end blood testing at the point-of-care: integration of robotic phlebotomy with downstream sample processing. *Technology* **6**, 59–66 (2018).
50. Drain, P. K. et al. Diagnostic point-of-care tests in resource-limited settings. *Lancet Infect. Dis.* **14**, 239–249 (2014).
51. Ronneberger, O., Fischer, P. & Brox, T. U-Net: convolutional networks for biomedical image segmentation. *Med. Image Comput. Comput. Interv.* **9351**, 234–241 (2015).
52. Ioffe, S. & Szegedy, C. Batch normalization: accelerating deep network training by reducing internal covariate shift. In *Proc. 32nd International Conference on Machine Learning* Vol. 37, 448–456 (JMLR, 2015).
53. He, K., Zhang, X., Ren, S. & Sun, J. Delving deep into rectifiers: surpassing human-level performance on ImageNet classification. In *Proc. 2015 IEEE International Conference on Computer Vision* 1026–1034 (IEEE, 2015).
54. He, K., Zhang, X., Ren, S. & Sun, J. Identity mappings in deep residual networks. In *Proc. 14th European Conference on Computer Vision* 630–645 (Springer, 2016).
55. Rumelhart, D. E., Hinton, G. E. & Williams, R. J. Learning representations by back-propagating errors. *Nature* **323**, 533–536 (1986).
56. Shi, X. et al. Convolutional LSTM networks: a machine learning approach for precipitation nowcasting. In *Proc. 28th International Conference on Neural Information Processing Systems* 802–810 (MIT Press, 2015).
57. Cho, K. et al. Learning phrase representations using RNN encoder–decoder for statistical machine translation. In *Proc. 2014 Conference on Empirical Methods in Natural Language Processing* 1724–1734 (ACL, 2014).
58. Chung, J., Çağlar, G., Cho, K. & Bengio, Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. Preprint at <https://arxiv.org/abs/1412.3555> (2014).
59. Chen, A. I., Balter, M. L., Maguire, T. J. & Yarmush, M. L. 3D near infrared and ultrasound imaging of peripheral blood vessels for real-time localization and needle guidance. In *Medical Image Computing and Computer-Assisted Interventions* Vol. 9902, 130–137 (Springer, 2016).
60. Zhao, H., Gallo, O., Frosio, I. & Kautz, J. Loss functions for image restoration with neural networks. *IEEE Trans. Comput. Imaging* **3**, 47–57 (2017).
61. Sudre, C. H., Li, W., Vercauteren, T., Ourselin, S. & Jorge Cardoso, M. in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support* (eds Cardoso M. et al.) 240–248 (Springer, 2017).
62. Chambolle, A., Caselles, V., Novaga, M., Cremers, D. & Pock, T. in *Theoretical Foundations and Numerical Methods for Sparse Recovery* (ed. Fornasier, M) 263–340 (2010).
63. Říha, K. *Artery Databases* (Brno University of Technology, 2014); [http://splab.cz/wp-content/uploads/2014/05/ARTERY\\_TRANSVERSAL.zip](http://splab.cz/wp-content/uploads/2014/05/ARTERY_TRANSVERSAL.zip)
64. Zukal, M., Beneš, R., Číka, I. P. & Říha, K. *Ultrasound Image Database* (Brno University of Technology, 2013); [http://splab.cz/wp-content/uploads/2013/11/us\\_images.zip](http://splab.cz/wp-content/uploads/2013/11/us_images.zip)
65. Le Guennec, A., Malinowski, S. & Tavenard, R. Data augmentation for time series classification using convolutional neural networks. Preprint at <https://halshs.archives-ouvertes.fr/halshs-01357973> (2016).
66. Kingma, D. P. & Ba, J. L. Adam: a method for stochastic optimization. In *Proc. 3rd International Conference on Learning Representations* (ICLR, 2015).
67. Ng, A. Y. Feature selection, L1 vs. L2 regularization, and rotational invariance. In *Proc. 21st International Conference on Machine Learning* 78–85 (ACM, 2004).
68. Abadi, M. et al. TensorFlow: a system for large-scale machine learning. In *Proc. 12th USENIX Conference on Operating Systems Design and Implementation* Vol. 16, 265–283 (USENIX, 2016).
69. Liu, B., Zhang, F. & Qu, X. A method for improving the pose accuracy of a robot manipulator based on multi-sensor combined measurement and data fusion. *Sensors* **15**, 7933–7952 (2015).
70. Bradski, G. The OpenCV Library. *Dr Dobbs J. Softw. Tools* **120**, 122–125 (2000).
71. Rusu, R. B. & Cousins, S. 3D is here: Point Cloud Library (PCL). In *Proc. 2011 IEEE International Conference on Robotics and Automation* 1–4 (IEEE, 2011).
72. Yoo, T. S. et al. Engineering and algorithm design for an image processing API: a technical report on ITK—The Insight Toolkit. *Stud. Health Technol. Inform.* **85**, 586–592 (2002).
73. Myronenko, A. & Song, X. Point set registration: coherent point drifts. *IEEE Trans. Pattern Anal. Mach. Intell.* **32**, 2262–2275 (2010).
74. Gunst, R. F. & Mason, R. L. Fractional factorial design. *WIREs Comput. Stat.* **1**, 234–244 (2009).

## Acknowledgements

We thank J. Leiphemer and N. DeMaio for their assistance and support in designing and implementing the device, E. Pantin and A. Davidovich for support in the human imaging studies and phantom studies, including review of imaging data and overall clinical guidance, E. Yurkow, D. Adler, M. Lo and G. Yarmush for assistance in the animal studies, and B. Lee for code used in our deep learning approach. Research reported in this publication was supported by the National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health under awards R01-EB020036 and T32-GM008339. This work was also supported by a National Institutes of Health Ruth L. Kirschstein Fellowship F31-EB018191 awarded to A.I.C. and a National Science Foundation Fellowship DGE-0937373 awarded to M.L.B. The authors acknowledge additional support from the Rutgers University School of Engineering, Rutgers University Department of Biomedical Engineering and the Robert Wood Johnson University Hospital.

## Author contributions

A.I.C. and M.L.B. designed the system, developed the algorithms and annotation software, and implemented the software and hardware. A.I.C. led execution of the imaging, in vitro and in vivo studies and analysed the primary data. M.L.Y. and T.J.M. provided the general direction for the project and provided valuable comments on the system design and manuscript. Correspondence and requests for materials should be addressed to A.I.C.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s42256-020-0148-7>.

**Correspondence and requests for materials** should be addressed to A.I.C.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020