


# Towards expert-level autonomous carotid ultrasonography with large-scale learning-based robotic system

Received: 28 October 2024

Accepted: 30 July 2025

Published online: 23 August 2025

 Check for updates

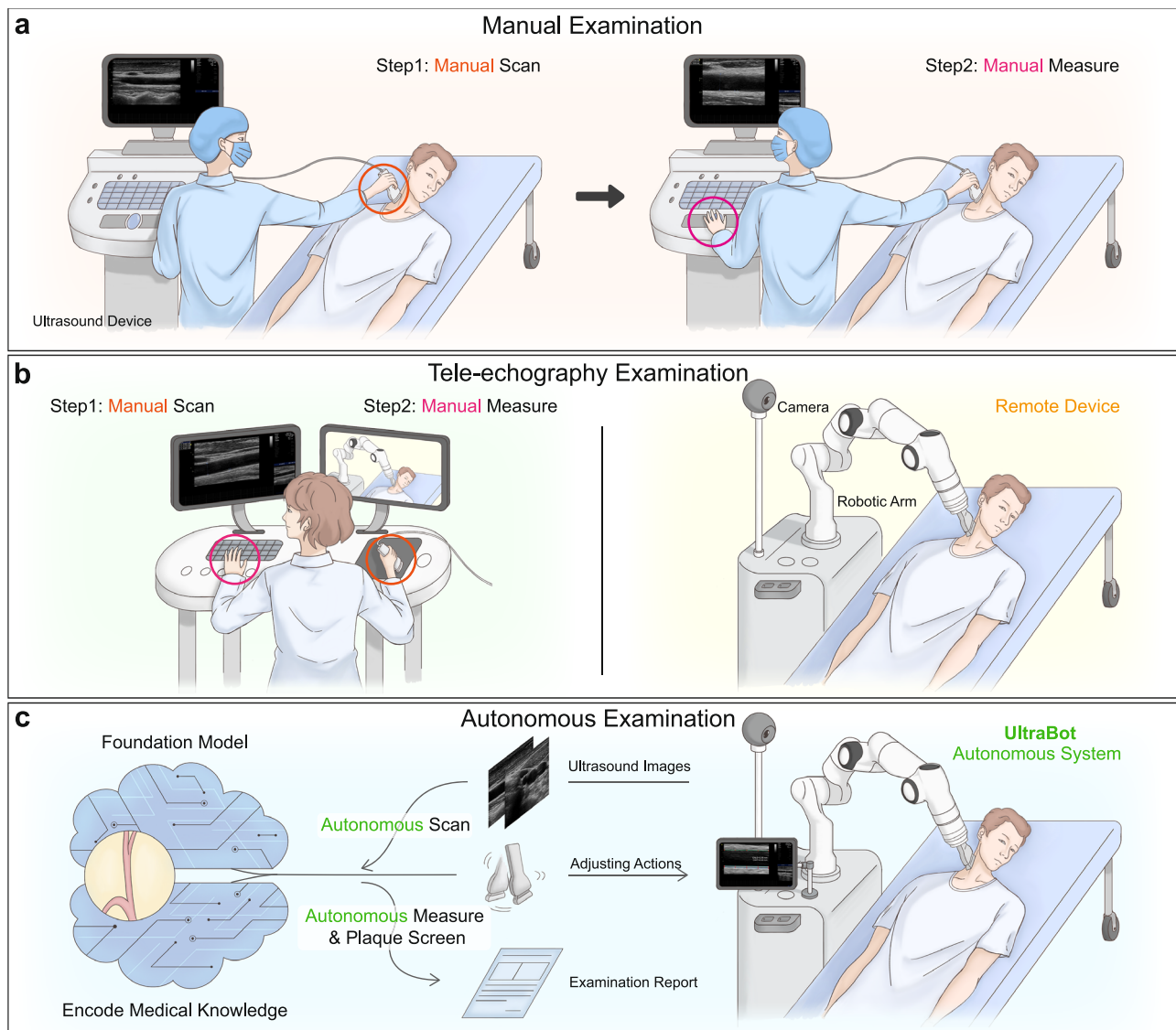
Haojun Jiang <sup>1,2,8</sup>, Andrew Zhao <sup>1,2,8</sup>, Qian Yang <sup>3,8</sup>, Xiangjie Yan <sup>1,2</sup>, Teng Wang <sup>1,2</sup>, Yulin Wang <sup>1,2</sup>, Ning Jia <sup>4</sup>, Jiangshan Wang <sup>5</sup>, Guokun Wu <sup>5</sup>, Yang Yue <sup>1,2</sup>, Shaqi Luo<sup>6</sup>, Huanqian Wang <sup>1,2</sup>, Ling Ren<sup>7</sup>, Siming Chen<sup>7</sup>, Pan Liu <sup>7</sup>, Guocai Yao <sup>6</sup>, Wenming Yang<sup>5</sup>, Shiji Song<sup>1,2</sup>, Xiang Li <sup>1,2</sup>, Kunlun He <sup>7</sup>  & Gao Huang <sup>1,2</sup> 

Carotid ultrasound requires skilled operators due to small vessel dimensions and high anatomical variability, exacerbating sonographer shortages and diagnostic inconsistencies. Prior automation attempts, including rule-based approaches with manual heuristics and reinforcement learning trained in simulated environments, demonstrate limited generalizability and fail to complete real-world clinical workflows. Here, we present UltraBot, a fully learning-based autonomous carotid ultrasound robot, achieving human-expert-level performance through four innovations: (1) A unified imitation learning framework for acquiring anatomical knowledge and scanning operational skills; (2) A large-scale expert demonstration dataset (247,000 samples, 100 × scale-up), enabling embodied foundation models with strong generalization; (3) A comprehensive scanning protocol ensuring full anatomical coverage for biometric measurement and plaque screening; (4) The clinical-oriented validation showing over 90% success rates, expert-level accuracy, up to 5.5 × higher reproducibility across diverse unseen populations. Overall, we show that large-scale deep learning offers a promising pathway toward autonomous, high-precision ultrasonography in clinical practice.

Ultrasonography is an integral medical imaging technique in contemporary medical diagnostics, utilized from the fetal stage<sup>1,2</sup> throughout an individual's life. It enables health assessments of a wide range of organs<sup>1–8</sup>, including the carotid artery<sup>4,5,9</sup>, heart<sup>3,6</sup>, liver<sup>7,8</sup>, among others. Compared to X-ray imaging, ultrasound boasts several predominant advantages: it offers real-time, dynamic visualization of the organs and tissues in a radiation-free and cost-effective manner, offering rich information for clinical diagnostics.

However, unlike other medical imaging techniques, ultrasound examination significantly relies on manual operation (Fig. 1a, b). Sonographers require close coordination of eye, hand, and brain for real-time decision-making to adjust the transducer's pose, with a nontrivial challenge in adapting scanning strategies to individual differences on a case-by-case basis. This heavy reliance on sonographers' experience diminishes the standardization and accuracy of examinations, leading to high variability in inter-sonographer results<sup>10–12</sup>,

<sup>1</sup>Department of Automation, Tsinghua University, Beijing, China. <sup>2</sup>Beijing National Research Center for Information Science and Technology, Tsinghua University, Beijing, China. <sup>3</sup>Air Force Medical Center, Beijing, China. <sup>4</sup>LeadVision Ltd, Beijing, China. <sup>5</sup>Shenzhen International Graduate School, Tsinghua University, Shenzhen, China. <sup>6</sup>Beijing Academy of Artificial Intelligence, Beijing, China. <sup>7</sup>Chinese PLA General Hospital, Beijing, China. <sup>8</sup>These authors contributed equally: Haojun Jiang, Andrew Zhao, Qian Yang. ✉e-mail: [kunlunhe@plagh.org](mailto:kunlunhe@plagh.org); [gaohuang@tsinghua.edu.cn](mailto:gaohuang@tsinghua.edu.cn)



**Fig. 1 | Comparison of manual, tele-echography, autonomous robotic ultrasound examination.** **a** Manual ultrasonography: sonographers conduct ultrasound scans using a handheld transducer and manually measure biometrics. **b** Tele-echography: sonographers operate the transducer on a simulator, and then the adjustments are executed by a distant robotic arm holding the transducer,

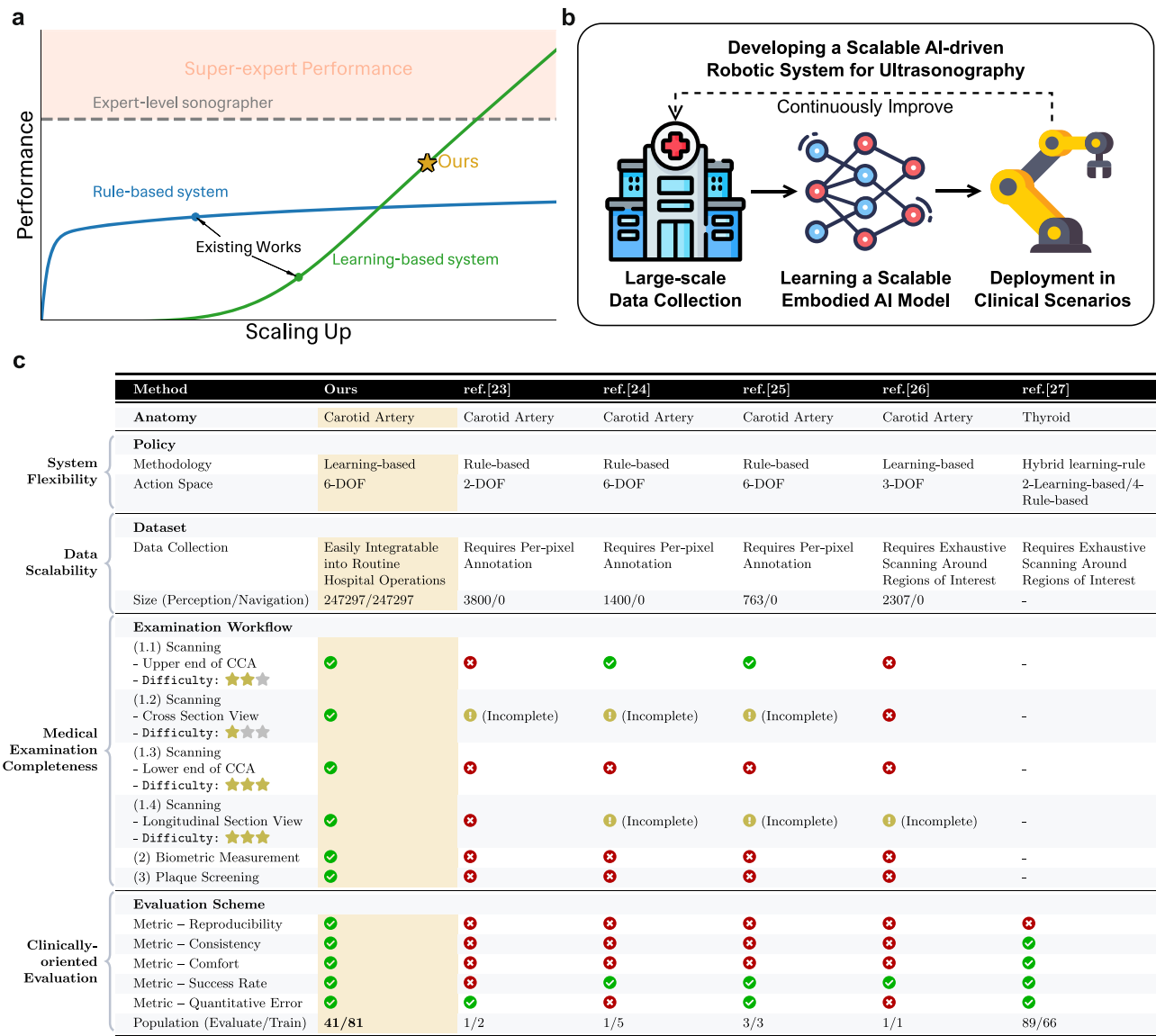
mirroring the sonographer's actions. They also manually measure biometrics on the remote control device. **c** Autonomous robotic ultrasonography: empowered by the foundation model, the robotic system automatically performs scanning, biometric measurements, and plaque screening, demonstrating significant clinical potential.

particularly among less experienced sonographers or trainees. Moreover, the complexity of manual operation makes training sonographers a time-intensive endeavor<sup>13</sup>, contributing to a significant shortage of these professionals<sup>10,14</sup>. Additionally, manual operation requiring close contact with patients can increase the risk of infection for practitioners during epidemic periods<sup>15,16</sup>.

On the flip side, medical robots with a high level of autonomy present a potential solution to mitigate the dependence on sonographers' experience and availability, thereby enhancing the examination process. In this paper, we investigate how to develop an expert-level, fully autonomous robot, which dynamically analyzes the ultrasound signals collected from patients, adjusts the probe's moving trajectories and poses in real time, and accomplishes scanning and measuring tasks in real clinician scenarios (Fig. 1c). In particular, we use carotid artery ultrasonography (Fig. 1a) as a case study, which is a common, important approach for assessing cardiovascular diseases. For example, it can be conveniently employed for the effective detection of stenosis and atherosclerotic plaques<sup>17–19</sup>. These conditions

are notable risk factors for cardiovascular diseases, which have impacted 422.7 million individuals worldwide, resulting in 17.9 million fatalities and constituting 31% of the total global deaths in 2015<sup>20–22</sup>.

Currently, most medical ultrasound robotic systems (Fig. 2c) adopt rule-based decision strategies, depending on a predefined set of rules to guide the decision-making process<sup>23–25</sup>. For example, in work<sup>23</sup>, the core scanning carotid trajectory relied on manually pre-defined paths, meaning all action decisions are predetermined. Meanwhile, works<sup>24,25</sup> implemented traditional visual servoing for carotid scanning, where action decisions were made based on rule sets derived from observed image feature changes following each executed movement. (See supplementary material for details.) However, it is inherently difficult to define a set of sufficiently generalizable rules applicable to complex clinical environments, e.g., adapting to the large variability of each individual's carotid artery in structure, morphology, and position. Hence, rule-based strategies are usually suitable only for a limited range of scenarios, lacking the necessary ability to fulfill all ultrasound examination tasks or generalize to new individuals. In



**Fig. 2 | Comparison with existing works.** **a** Two main types of ultrasound robotic systems: rule-based and learning-based. As learning-based methods scale up with data and model size, they will show superior generalizability compared to rule-based methods, further possessing the potential to surpass human experts. **b** Our philosophy is to embrace the scaling law, involving the large-scale collection of expert data, training scalable neural networks, and future deployment in real clinical settings to establish a loop that enables continuous data and model scaling. **c** We compare our system with existing works across four critical aspects: system flexibility, data scalability, comprehensiveness of medical examinations, and

clinically-oriented evaluation. ☆ indicates task difficulty. The “Incomplete” label denotes scans covering only a partial carotid artery segment (between the internal/external and common carotid-subclavian junctions), rather than the full vessel. The abbreviation “CCA” stands for the common carotid artery, with its upper end marking the bifurcation into internal/external carotid arteries and its lower end indicating the junction with the subclavian artery. The quantitative error metric evaluates errors in at least one area: target segmentation or biometric measurement.

contrast to rule-based methods, researchers have recently begun to explore harnessing learnable neural networks as the controller of ultrasound robots<sup>26–28</sup>. For instance<sup>26</sup>, adopted a reinforcement learning approach for motion planning, but trained on simulated and simplified vascular virtual environments. Nevertheless, control strategies trained on simplified vascular fail to generalize to real human anatomical variations, limiting clinical applicability. Representative thyroid scanning work<sup>27</sup> adopted a hybrid learning-rule strategy, implementing learning-based methods for two degrees of freedom (DoF) while maintaining rule-based control for the remaining four. This hybrid strategy remains constrained by the limitations of rule-based methods, making it inadequate for handling complex anatomical variations like

those in carotid arteries. (See supplementary material for details.) Learning-based methods, in theory, have the potential of fitting functions with arbitrary complexities and therefore may address the issues of predefined rules (Fig. 2a). Nevertheless, current learning-based approaches have mainly been investigated in small-scale, experimental scenarios, with significant limitations in data, modeling, and validation. First, their data collection schemes tend to be labor-intensive, less realistic, and difficult to scale up (e.g., constructing a phantom or collecting all possible probe positions and poses for each person), yielding a limited number of learning samples, sometimes with training-test domain shifts (Fig. 2c, Data Scalability). Second, with the relatively small-scale training data, the action space is typically defined

with limited degrees of freedom (*e.g.*, translation or rotation in one direction) to avoid overfitting and ensure training stability (Fig. 2c, System Flexibility). Such simplifications often make it difficult for the system to handle complex situations in real world. Third, validation of the system is usually conducted only on a few individuals or even phantom, without considering clinically-oriented validations (Fig. 2c, Clinically-oriented Evaluation). This is insufficient to fully evaluate its applicability in clinical scenarios.

In recent years, the rise of large language models and multi-modal foundation models like GPT-4V<sup>29</sup> has demonstrated the scaling law of neural networks<sup>30,31</sup>: previously unobserved abilities can emerge with growing datasets, increases in model size, and advances in learning algorithms, such as the capabilities of handling more complex situations and generalizing to new, unseen samples without additional training. Inspired by this success, this paper explores the potential of scaling up a purely learning-based framework in autonomous robotic ultrasound examination. Through four key contributions—(1) more flexible modeling formulations, (2) utilization of large-scale real-world individual data, (3) a more comprehensive examination process, and (4) validation on a broader population with clinically oriented evaluation protocols—we demonstrate the powerful capabilities of a fully learning-driven ultrasound robot. Our work not only highlights the system's promising potential but also charts a viable path to bridge the gap between theoretical research and real-world clinical adoption.

At its core, our system frames autonomous scanning as an end-to-end vision-driven navigation task, where the transducer's movements are inferred from real-time ultrasound images, under the goal of obtaining high-quality diagnostic images. Specifically, we integrate perception and scanning action decision-making within a unified network, optimized through a deep imitation learning strategy<sup>32–34</sup> that eliminates the need for handcrafted rules<sup>23,24</sup> (Fig. 2c, System Flexibility). This network achieves 6-DoF probe pose control through a unified learning strategy governing all DoFs, resulting in a simplified yet efficient system architecture compared to prior work<sup>27</sup>. The 6-DoF configuration ensures effective handling of the carotid artery's complex anatomical structures, whereas systems<sup>23,26</sup> with limited degrees of freedom struggle to adapt to the demands of comprehensive ultrasound scanning. In terms of data, our philosophy is to embrace data scaling law, learning generalizable scanning strategies from extensive data (Fig. 2c, Data Scalability). Thus, we collected a large-scale dataset of expert demonstrations of carotid artery scans performed on real individuals, comprising 247,297 pairs of ultrasound images and corresponding scanning actions, encompassing a wide range of individual tissue structural variations likely to be encountered in the real world and corresponding expert adaptation actions. To the best of our knowledge, this dataset is 100 times larger than those used in previous works<sup>23–26</sup>. Notably, these expert demonstration data naturally exist in routine medical ultrasound examinations but were previously unrecorded. Since our data collection process eliminates the need for complex annotations (unlike prior works<sup>23–25</sup>), further scaling the dataset is entirely feasible. By training the neural network with deep imitation learning on this dataset, the model can acquire generalizable knowledge and skills for navigation, including anatomical knowledge, ultrasound image interpretation ability, and transducer operation skills.

Building upon our flexible modeling framework and large-scale data learning approach, our robotic system achieves unprecedented comprehensiveness in carotid artery scanning, enabling thorough vascular assessments (Fig. 2c, Medical Examination Completeness). Previous efforts<sup>23–26</sup> were constrained by rule-based strategies, limited degrees of freedom, and small-scale datasets, making comprehensive scanning unattainable. More importantly, we achieve a fully autonomous workflow integrating scanning, measurement, and plaque screening, overcoming the limitations of conventional approaches<sup>23–26,35,36</sup> that focused on isolated steps, thereby establishing

a crucial foundation for developing truly practical end-to-end autonomous ultrasound robotic systems.

Finally, our work conducts a clinically-oriented evaluation, while also scaling up the evaluation population by an order of magnitude compared to works<sup>23–26</sup> (Fig. 2c, Clinically-oriented Evaluation). This approach enables more accurate and clinically relevant validation, better reflecting the practical potential of autonomous ultrasound robots. Concretely, our robotic system achieves over 90% scanning success rate across a diverse population (Age: 19–70 years old; Body Mass Index (BMI): 16.5–30.8; Sex: female and male), confirming its strong generalization performance across anatomical variations, including successful scanning of patients with plaques. Meanwhile, existing carotid artery studies<sup>23–26</sup> suffer from critically limited validation cohorts (1–3 subjects), fundamentally constraining their ability to demonstrate robotic systems' clinical viability. Furthermore, we demonstrate the robotic system's capability for precise measurement of key anatomical structures that reflect the health status of carotid artery, *i.e.*, intima-media thickness and lumen diameter. Hypothesis testing shows that the robotic system's biometric measurements align with expert outcomes, evidenced by a *p*-value below 0.001. Notably, we present the validation of a robotic ultrasound system's reproducibility in biometric measurements, demonstrating superior performance across all four reproducibility metrics (with improvements up to 5.5 times)—marking a significant advancement that overcomes fundamental limitations of conventional manual scanning techniques. Moreover, the robotic system demonstrates automated precise plaque segmentation, achieving promising performance on real patients and expert-annotated datasets, thereby enabling pathological case detection capabilities. Overall, our empirical findings demonstrate the potential of this large-scale learning-based robotic system in clinical applications, marking a significant turning point in the development of AI-driven medical ultrasound robotics.

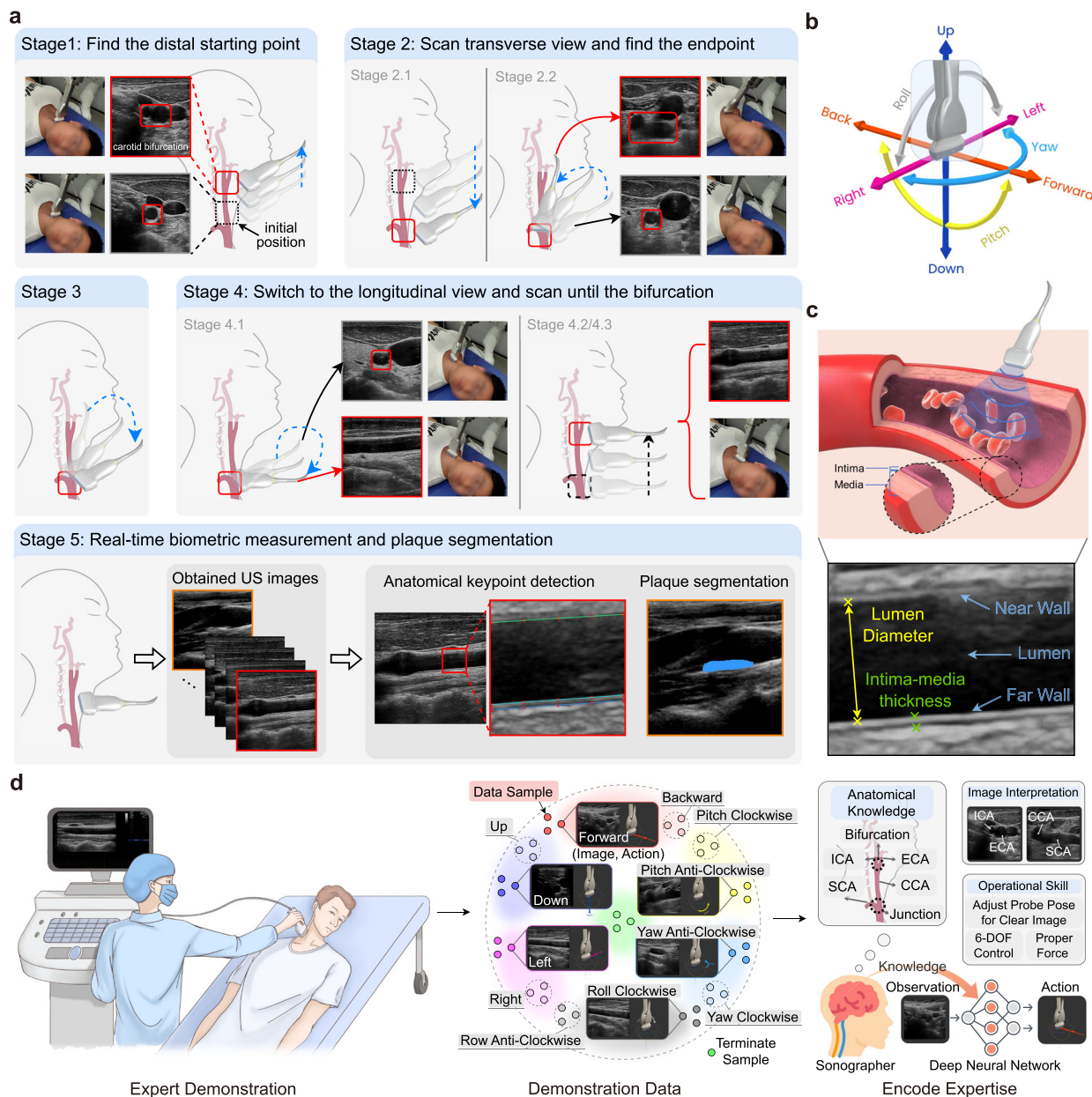
## Results

In this section, we initially present an overview of the functionalities and critical technologies of the autonomous ultrasound robotic system. We then conduct a comprehensive evaluation comparing its performance against experienced sonographers, assessing both result reproducibility and consistency. The system's robustness is further validated across diverse clinical variables including patient age, BMI, different ultrasound machines, and imaging parameters. Finally, we provide a validation of the algorithms for each component of the system using expert-annotated datasets.

### Autonomous carotid ultrasound compliant with medical standards

Clinical ultrasound examination is an intricate and demanding task that necessitates the meticulous coordination of a professional sonographer's hand, eye, and brain. In this article, we delve into this highly challenging issue and design an autonomous ultrasound robotic system specifically for clinical carotid artery examinations. The “hand” is represented by a 7 degrees of freedom (7-DOF) Franka Emika Panda robotic arm which holds an ultrasound transducer, the “eye” is facilitated through an ultrasound imaging system and an external camera, and the “brain” comprises a series of deep neural networks that encapsulate expert knowledge. In accordance with professional medical guidelines<sup>19,37,38</sup>, we categorize the robotic ultrasound examination into the following stages: scanning (Stages 1–4) and image analysis (Stage 5). Each stage is tailored to clinical examination requirements, with the goal of capturing high-quality imaging data, acquiring essential biometric measurements, and detecting the presence of plaques to assess carotid artery health (Fig. 3a). Furthermore, upon obtaining clear longitudinal images of the carotid artery, the ultrasound system's integrated Color Doppler and Pulse Doppler functions can be immediately activated to assess the subject's hemodynamic





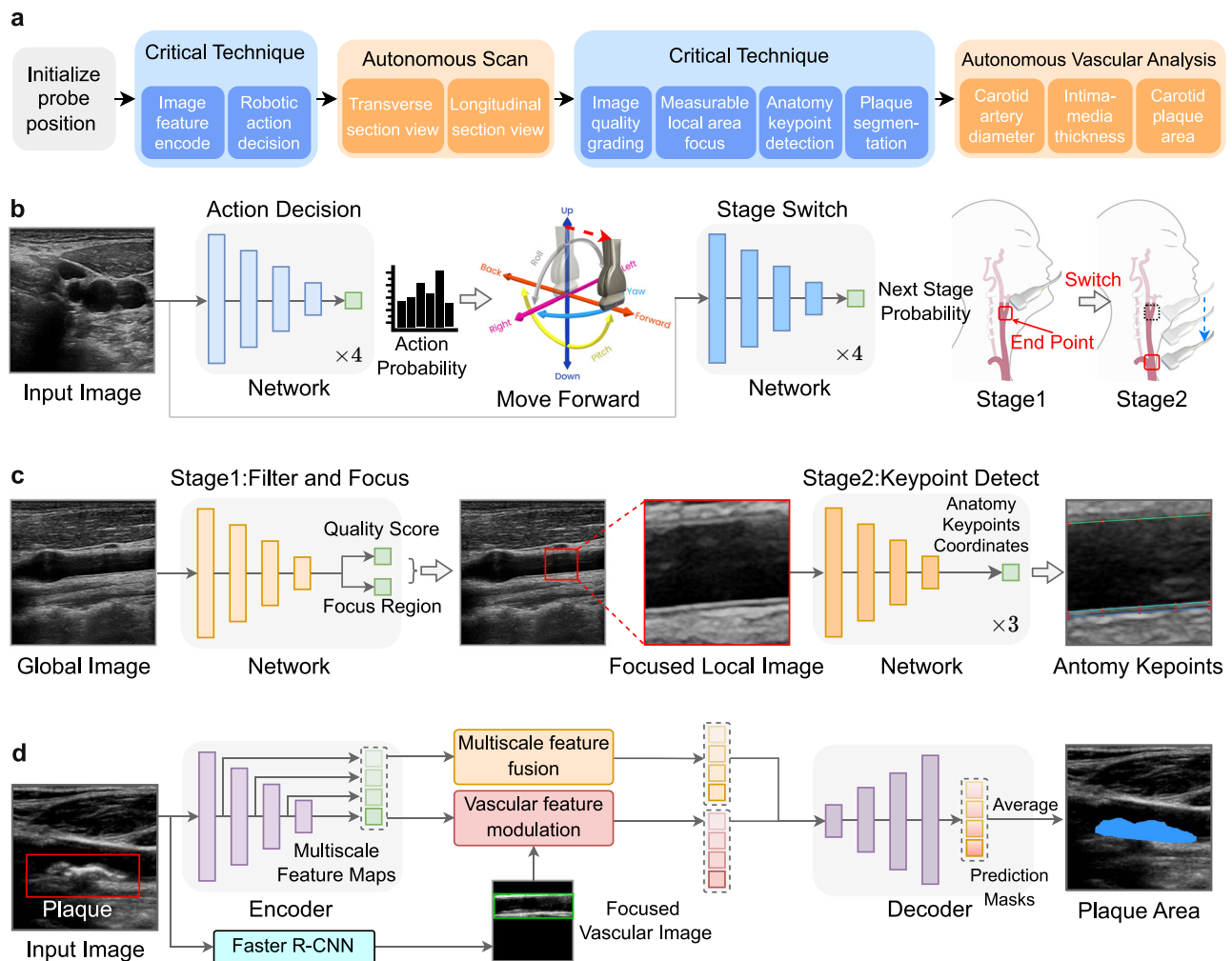
**Fig. 3 | Autonomous carotid artery ultrasound examination system workflow.** **a** Schematic diagram of each stage, including the relative position of the transducer to the carotid artery, the motion posture of the transducer, and the corresponding ultrasound images. Finally, assess the collected images' quality and select the appropriate ultrasound images for biometric measurement and plaque segmentation. **b** The coordinate system of the transducer and its action space, 6 degrees of freedom, 12 discrete actions. **c** Longitudinal section diagrams of the carotid artery, vascular wall structures, and the schematic diagram of biometrics. **d** Collecting

high-quality demonstration data (action-image pair) from experts, specifically how to adjust the transducer to obtain ultrasound images suitable for measurement or diagnosis. Subsequently, leveraging an imitation learning paradigm, we encapsulate experts' knowledge within deep neural networks to facilitate autonomous ultrasound scanning. The term "CCA", "SCA", "ICA", and "ECA" refers to the common carotid artery, subclavian artery, internal carotid artery, and external carotid artery, respectively.

profile. Live demonstrations can be seen in the Supplementary Video 1–3. The advantage of designing the robotic process based on medical rules is that it endows each stage with a clear medical purpose and significance, while also enhancing the standardization of ultrasonography.

- Stage 1 - Locate the distal starting point of the common carotid artery: Starting from the transverse view of the right common carotid artery, the ultrasound transducer moves up upward along the neck, keeping the carotid artery in the image, until it identifies the bifurcation of the internal and external carotid arteries and then proceeds to the next stage.

- Stage 2.1 - Transverse scanning of the common carotid artery: The ultrasound transducer glides along the right common carotid artery, keeping it centrally in view, until it reaches the point where it intersects with the subclavian artery.
- Stage 2.2 - Locate the proximal endpoint of the common carotid artery: The robot moves the ultrasound transducer out-of-plane, making accurate adjustments to clearly display the junction between the right common carotid and subclavian arteries, before advancing to the subsequent stage.
- Stage 3 - Reset the transducer's posture for the next stage: The ultrasound transducer primarily executes a "pitch anti-clockwise"



**Fig. 4 | Autonomous scanning and image analysis with deep neural networks.** **a** Technical workflow for autonomous carotid artery ultrasound examination system, including autonomous scan, biometric measurement, and plaque segmentation. **b** The scanning model architecture demonstrates the action decision module and stage transition module. **c** The biometric measurement model architecture

includes the image quality assessment module, measurable region focusing module and intima structure keypoint detection module. **d** The plaque segmentation model architecture employs the vessel region focusing module and multi-scale feature fusion module.

action until the subclavian artery disappears from the image, then moves on to the next stage.

- Stage 4.1 - Switch from transverse to longitudinal view: The system adjusts the posture of the transducer until the longitudinal view of the right common carotid artery is clearly presented in the image.
- Stage 4.2 - Longitudinal scanning of the common carotid artery: Maintaining the longitudinal view, the ultrasound transducer move towards the bifurcation of the internal and external carotid arteries.
- Stage 4.3 - Termination of the scan: The scan concludes once the carotid bifurcation is visible in the longitudinal view.
- Stage 5 - Image analysis: During scanning, the algorithm automatically calculates biological parameters and detects plaque presence using stage 4-acquired images. The Color and Pulse Doppler functions of the ultrasound machine can then be employed to assess the subject's hemodynamic status.

A suite of deep learning-based technologies facilitates the autonomous operations mentioned above. These include the encoding of features in ultrasound images, decision-making for robotic actions, evaluation of image quality, concentration on specific local areas, identification of key anatomical points, and detects plaque presence and outlines contours (Fig. 4a).

**Purely learning-based navigation framework:** Autonomous ultrasound scanning is analogous to an autonomous driving task performed on the human body's surface. This difficult task requires generalizable perception, decision-making, and control technologies to navigate the transducer and conduct a safe, effective, and autonomous scan. For successful execution, the entire system must possess the following knowledge and capabilities (Fig. 3d right): (1) A thorough understanding of the carotid artery's anatomical structure, including the pathway of the carotid artery, the structure of bifurcations, and corresponding blood vessels, analogous to a "map" in autonomous driving; (2) The capability to interpret ultrasound images and spatial imagination skills, requiring the identification of structures on ultrasound images and mapping these two-dimensional structures onto a three-dimensional anatomical framework to locate the current position of the transducer; (3) Proficiency in operating the transducer, encompassing knowledge of how to change the probe's pose to obtain clear imaging and how to apply appropriate force. As discussed in Sec. 5, most previous works<sup>23–25</sup> adopt rule-based decision strategies, depending on a predefined set of rules to guide the decision-making process. Nevertheless, pre-determined rules often fail to encapsulate the full scope of the aforementioned knowledge and capabilities with sufficient flexibility, frequently encountering difficulties in handling complex individual

differences and exhibiting limited generalizability when applied to new individuals in clinical settings.

Diverging from prior approaches, we introduce a purely learning-based navigation framework. Our core concept encompasses two main aspects: first, integrating perception and decision-making into a unified network, optimizing it directly based on the final navigation goal; and second, entirely learning generalizable skills and knowledge from extensive expert data. To achieve data-driven autonomous scanning, we build a large-scale carotid artery scanning expert demonstration dataset, containing 247K pairs of ultrasound images and corresponding scanning actions. Specifically, professionals are guided to execute scans following the predetermined procedure, recording the expert maneuvers as 12 discrete actions across six degrees of freedom (Fig. 3b), with an additional action denoting the conclusion of each stage. In this dataset, the ultrasound imaging data include variations in tissue structures across different individuals. Additionally, the mapping relationship from ultrasound images to scanning actions implicitly contains anatomical knowledge, ultrasound image interpretation ability, and transducer operation skills. Inspired by recent advancements in imitation learning<sup>32–34</sup>, which is a method well-suited for modeling complex tasks, we adopt this strategy to embed the required knowledge into our unified network (Fig. 3d). After optimization based on the ultimate navigation goal, the action decision network can autonomously decide the subsequent action based on the current ultrasound image. Meanwhile, a stage switch network is trained to identify anatomical landmarks indicating the end of a stage and instruct the system to proceed to the next stage. Additionally, from Stage 1 to 4, due to the different tasks at each stage, we trained four separate networks (Fig. 4b). Please refer to Sec. 5 for implementation details.

**Interpretable autonomous biometric measurement:** After conducting a scan of the carotid artery, sonographers usually measure the lumen diameter and intima-media thickness on the longitudinal section images of the carotid artery (Fig. 3c). These two parameters are instrumental in evaluating the health status of the carotid artery, providing insights into potential risks associated with cardiovascular diseases. The typical procedure involves sonographers selecting images that clearly depict vascular wall structures. They then zoom into a localized area for measurement, manually operating the device's cursor to measure from point to point. In this paper, we present an automated measurement process designed to emulate expert behavior. To ensure that sonographers can ascertain the acceptability of the final results, we incorporate interpretable strategies in the final image analysis stage (Fig. 3a stage-5 and Fig. 4c). Interpretability<sup>39,40</sup> is of paramount importance in modern medical settings, as sonographers need to confirm the final outcomes. Furthermore, aiding sonographers in quickly and accurately comprehending and assessing the results is a pivotal step towards enhancing the overall efficacy of medical processes. Specifically, we predict three groups of keypoints, uniformly distributed along the X-axis, on the localized image. These points correspond to the upper intima (near wall), lower intima (far wall), and lower media (far wall) of the artery wall, respectively. With these points identified, we calculate the distances between corresponding points, adjusting for the artery's slope, to derive the final measurements. Interpretability is inherently embedded in this modeling approach, as sonographers can assess the credibility of the final output based on the visualized positions of key points. Please refer to Sec. 5 for implementation details.

**Precise autonomous plaque segmentation:** Carotid plaque represents one of the most prevalent pathological changes in the carotid artery system, serving as a well-established biomarker for elevated cardiovascular risk. Early identification and characterization of these atherosclerotic lesions through advanced imaging techniques enables timely intervention, which may effectively halt disease progression and prevent subsequent vascular complications. To

address this, we propose an innovative plaque segmentation algorithm based on a vessel region-focused mechanism to identify plaques in the longitudinal view (Fig. 3a stage-5 and Fig. 4d). The core idea is to prioritize vascular regions to reduce background interference, since plaques typically exist within blood vessels. Specifically, we train a Faster R-CNN detector<sup>41</sup> to accurately locate vascular regions, whose features are then used to modulate the encoder's multi-scale features. This approach enables the model to concentrate more effectively on vascular characteristics while suppressing irrelevant background information. Simultaneously, we develop a multi-scale feature fusion module that effectively combines high-level semantic information with low-level contour details. The high-level features provide rich contextual understanding, while the low-level features retain fine spatial details. This fusion strategy ensures that both semantic and structural information are leveraged synergistically to improve overall segmentation quality. Finally, the decoder integrates both modulated vessel features and fused multi-scale features for precise plaque region prediction. Please refer to Sec. 5 for implementation details.

### System-level clinical evaluation on human subjects

Equipped with the aforementioned key components, the autonomous system is capable of executing the automated process of carotid artery ultrasound examination, thereby providing crucial biometric data. During our study, we recruited 122 volunteers, comprising 81 individuals (61 males and 20 females) for the training and validation set, and 41 (26 males and 15 females) for the test set. We explained the full process to all participants, and each signed an informed consent form (including their understanding of publishing information in the journal). We built a large-scale dataset for learning intelligent scanning strategies, comprising 247,297 pairs of ultrasound images and corresponding expert operational data collected from 81 volunteers. After training, in order to ascertain the performance of this system in real clinical environments, we organized 41 previously unseen volunteers to undergo autonomous carotid ultrasound examinations. It is noteworthy that the data from these 41 volunteers were not included in the training set. This is an important setting for evaluating the real-world generalizability of our system, *i.e.*, in real clinical settings, ultrasound examinations are always performed on individuals not previously encountered. In this test population, the oldest participant was 70 years old, with 7 subjects over 60 (6 subjects exhibiting plaques), 7 aged between 45 and 60, and the remaining under 45. For detailed patients' pathological information, please refer to Supplementary Table 1. Moreover, these 41 volunteers display a wide range of physiques. Their heights vary between 1.55 to 1.90 meters (mean=1.72, sd=0.09), weights range from 46.0 to 100.0 kilograms (mean=65.0, sd=12.1), and BMI spans from 16.5 to 30.8 (mean=22.1, sd=3.2), with 12 subjects BMI < 20 and 13 subjects BMI ≥ 24. These variations in body size, which contribute to anatomical differences, invariably escalate the scanning complexity. During the validation experiments, a subset of volunteers (n=21) underwent three scans and measurements carried out by the autonomous system, in addition to one scan and measurement performed by each of the three sonographers. The results of these measurements, obtained from both groups, were subsequently compared and analyzed. The analysis extensively verified the autonomous system, evaluating it on the grounds of reproducibility, consistency, generalizability, efficiency, and comfort. Additionally, 15 older volunteers participated in age-robustness testing of the scanning algorithm, while 5 younger volunteers were included to validate the algorithm's robustness to imaging parameters and ultrasound machine variations. It's worth mentioning that, to the best of our knowledge, our investigation is pioneering in conducting a comprehensive assessment of the system, focusing on its practical medical value, which underscores the potential of such systems to evolve into real clinical applications. Contrarily, previous studies<sup>23,24,42–46</sup> have



mainly evaluated such systems from a technical standpoint, only emphasizing aspects like operational precision and success rate.

Autonomous system possesses superior result reproducibility: Ultrasound examinations, heavily reliant on manual experience and operation, often yield significant variations in measurement results among different sonographers in practice. Severe variations can lead to incorrect diagnostic outcomes, making reproducibility a crucial metric for autonomous ultrasound systems. To assess the system's reproducibility, we compare the results of multiple measurements of carotid artery lumen diameter (CALD), and carotid intima-media thickness (CINT) by the autonomous system on the same individual with those performed by different ultrasound sonographers on the same person. The comparison is drawn across four metrics, following the work cited<sup>47,48</sup>, namely, spearman's correlation coefficient (SCC), intra-class correlation coefficient (ICC), coefficient of variation (CV), and mean absolute difference (MAD) (Fig. 5b). A larger value in the first two metrics signifies better reproducibility, while a smaller value in the last two metrics is preferable. As depicted in Fig. 5b, the autonomous system surpasses the sonographer in all four indicators, for both CINT and CALD. Specifically, for CINT, there is a notable improvement where the autonomous system amplifies the Spearman's correlation coefficient by 5.5 times and reduces the coefficient of variation and mean absolute difference by 2.4 and 2.8 times, respectively. For CALD, the autonomous system reduces the coefficient of variation and the mean absolute difference by 3.1 and 2.3 times, respectively. This is noteworthy as measuring CINT presents a greater challenge due to the necessity for a more precise identification of the intima-media boundary.

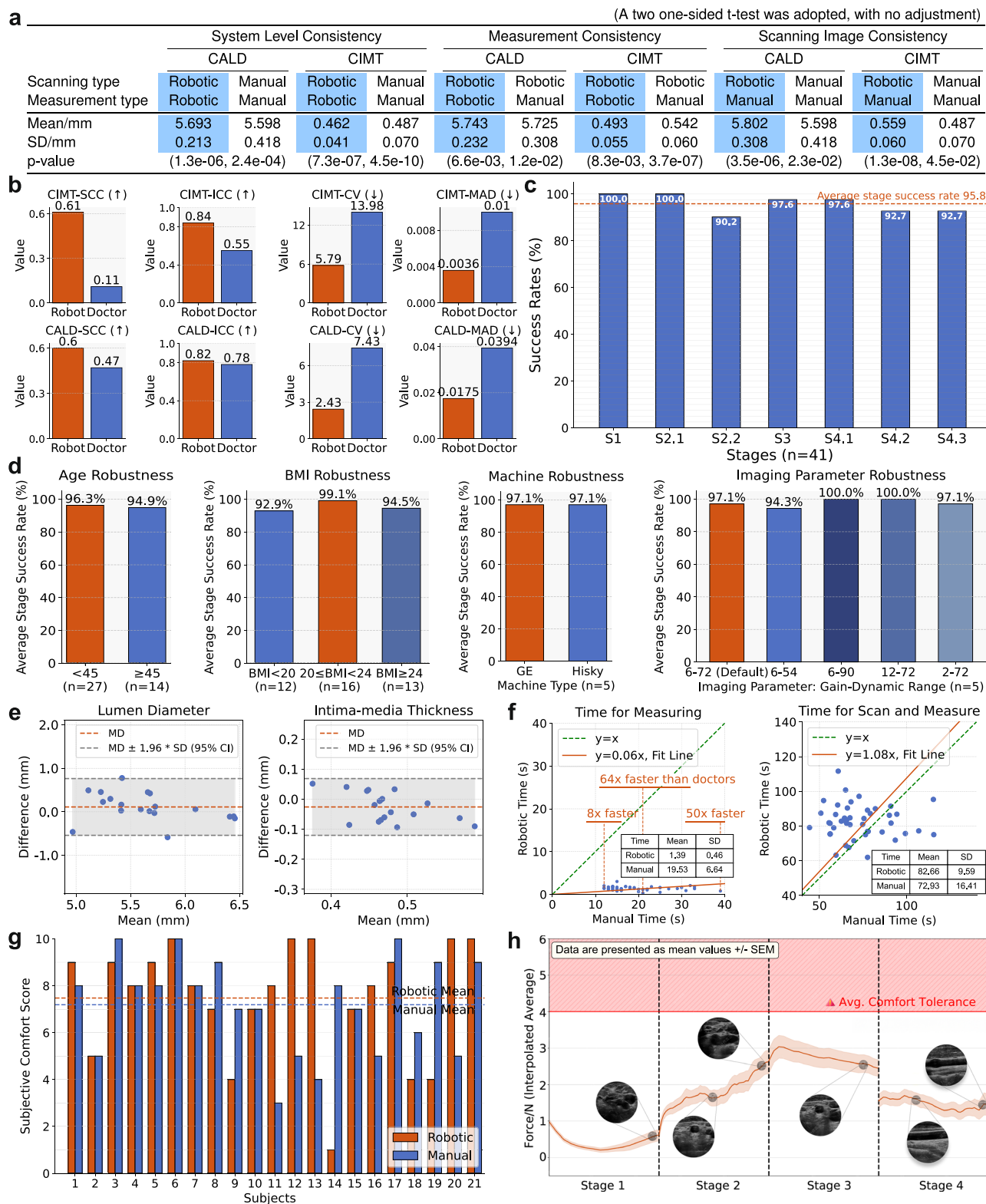
High consistency between autonomous and manual biometric measurements: The accuracy of an autonomous system's biometric measurement (CALD and CINT) is a critical factor when evaluating its potential deployment in real-world clinical settings. To verify this, we conduct an equivalence test<sup>49</sup> to compare the robotic system's results with professional sonographers' measurements. This two-sample equivalence test aims to determine whether the means of both populations can be considered statistically equivalent. Here, "equivalence" means the difference between the two group means falls within a predefined acceptable range, known as the equivalence margin. For CALD, we used the standard deviation among measurements taken by three senior sonographers. Since ultrasound measurements lack an absolute ground truth, the standard deviation among expert measurements reflects an acceptable range of variability. For CINT, we set the equivalence margin at 0.1 mm, which corresponds to the smallest measurable unit typically achievable by ultrasound devices, (e.g., General Electric, Vivid E7). In other words, if there is a discrepancy between the measurements of two sonographers, the smallest possible difference would be 0.1 mm, making this a reasonable tolerance threshold.

Specifically, we employ a two one-sided t-test (TOST) approach<sup>49</sup> to comprehensively validate consistency across three key dimensions: system-level consistency, measurement consistency, and image quality consistency (Fig. 5a). For system-level consistency testing, one set of data (CALD and CINT) is obtained through robotic autonomous scanning and measurement, while the other set is manually scanned and measured by three senior sonographers on the same group of subjects. Using the expert-obtained measurement data, we derived the inter-observer standard deviation of CALD measurements (0.418 mm) to establish the equivalence margin. Subsequently, we conducted the two one-sided test (TOST) procedure for formal equivalence testing. If both one-sided p-values were below 0.05, the two datasets were considered statistically equivalent. As shown in Fig. 5a, both one-sided p-values were below 0.05, demonstrating statistical equivalence between the robotic system results and sonographer-acquired measurements. For measurement consistency testing, On the same subjects, we had sonographers select one image from the robotic scans for

manual measurement, while simultaneously applying the our deep learning measurement algorithm to the same image, thereby obtaining two sets of biometric data. To assess measurement consistency, on the same subjects, sonographers select one image from the robotic scans for manual measurement while our deep learning algorithm simultaneously processes the same image, generating two comparable sets of biometric data from each subject. Then, we employ a TOST for equivalence testing, with predefined equivalence margins of 0.308 mm for CALD and 0.1 mm for CINT. The results again demonstrate that the measurements from the our deep learning algorithm were highly consistent with those of the sonographers. For image consistency evaluation within the same subject cohort, we compare two sets of data: robotic system-acquired images versus sonographer-acquired images, with all measurements performed by sonographers. Following the same analytical methodology, we perform equivalence testing using the TOST, applying equivalence margins of 0.418 mm for CALD and 0.1 mm for CINT. The results clearly indicate that the image quality obtained through robotic acquisition meets clinical measurement requirements Furthermore, we employ the Bland-Altman method to assess system-level consistency (Fig. 5e). The results show that the differences between the autonomous system and sonographer measurements consistently fall within the 95% confidence interval, demonstrating good consistency. Given that the quality of ultrasound images serves as the baseline for accurate measurements, the aforementioned results also validate the system's ability to deliver high-quality imaging outcomes.

Autonomous system exhibits good generalizability: In routine ultrasound examinations, sonographers frequently encounter new patients, each with unique anatomical variations. Notably, key vascular parameters such as carotid artery length, trajectory, and bifurcation location exhibit significant inter-individual variability. Furthermore, elderly individuals often present with atherosclerotic plaques that vary in size and spatial distribution. These inherent variations pose substantial challenges to autonomous ultrasound systems, demanding robust generalization capabilities. To evaluate the system's generalization capability, we recruited a cohort of 41 previously unexamined volunteers and assessed the success rate of autonomous scanning. The participant pool demonstrates substantial demographic and physiological diversity, including 7 subjects over 60 years old (maximum age 70), 6 patients with detectable atherosclerotic plaques, and 14 individuals above 45 years, as well as both underweight (minimal BMI 16.5) and obese (maximal BMI 30.8) cases. This heterogeneous population with its wide spectrum of anatomical variations provides a rigorous test for assessing the system's generalization performance. As shown in Fig. 5c, each participant underwent a maximum of three trial runs, with success rates exceeding 90.2% at all stages and reaching an average stage success rate of 95.8%. For failure cases, please refer to Supplementary Fig. 1 and the corresponding explanation in supplementary. Moreover, as shown in Fig. 5d, our system demonstrates robust generalization capabilities across elderly populations, successfully completing scans even for individuals with existing plaques while maintaining clear visualization of the lesions. We provide three scanning videos (Supplementary Video 1–3) demonstrating successful examinations in elderly subjects with plaques. Furthermore, despite the known challenges that varying fat distributions pose for ultrasound imaging, our model maintains high success rates across both underweight and overweight BMI groups (ranging from 16.5 to 30.8), further validating its superior generalization performance across diverse anatomical variations. Given that the expert scanning demonstration data (training data) was collected solely from 81 distinct individuals, this success rate is indeed promising, highlighting the potential of imitation learning in autonomous ultrasound scanning tasks. In the future, in accordance with scaling laws, we anticipate continued improvement in scanning success rates as more training data becomes available.





**Fig. 5 | Autonomous system evaluation results on unseen human subjects.**

**a** Hypothesis test of results consistency between of autonomous system and sonographers. **b** Testing the reproducibility of measurements of the CINT (first row) and CALD (second row) between the robot system and sonographers, evaluated by the SCC, ICC, CV, and MAD. **c** Success rate of the robot system on volunteers. **d** Robustness of the robotic system across variations in age, BMI, machine type, and imaging parameters. **e** Bland-Altman plot assessing the

consistency between lumen diameter and intima-media thickness measurements from the robot system and sonographers. **f** Comparison of the total time taken for scanning and measurement by the robot system and sonographers, as well as a comparison of just the measurement time. **g** Volunteers' subjective comfort perceptions under the operation of the robot system and sonographers. **h** Contact force in the Z-axis (up and down) direction between the transducer and the human neck during the scanning process.

Beyond the challenges posed by individual structural variations, sonographers also adjust imaging parameters on the ultrasound device based on case-specific conditions in clinical practice. We recruited five hold-out volunteers and systematically varied two key parameters, Gain (G) and Dynamic Range (DR), to evaluate the model's effectiveness. Regarding the effects of G and DR on imaging, please refer to Supplementary Fig. 2. As shown in Fig. 5d, despite changes in imaging parameters, the model still demonstrates high robustness and good success rate. This indicates that the model has learned semantic features from the large-scale data, specifically the characteristics of the carotid artery, rather than overfitting to low-level features such as image texture.

Furthermore, the clinical environment presents the additional challenge of adapting to diverse ultrasound devices. Specifically, we validate our model's scanning capability on an EQTouch ultrasound device (manufactured by Hisky Medical, Wuxi, China) using a linear probe (L15-4, Hisky). We conduct the experiments on the same five participants as those in the imaging parameter robustness study. As shown in Fig. 5d, our deep learning model made accurate scanning decisions and achieved a high success rate on the new machine. Although Supplementary Fig. 3 demonstrates low-level imaging differences between ultrasound devices, our model exhibits robust performance, indicating that its decision-making relies not on low-level image features but rather on semantic-level ones, thereby demonstrating strong generalization capability.

Existing studies<sup>24–26</sup> have reported scanning success rates for few operational stages, we comprehensively present these comparative metrics in Supplementary Table 3 for intuitive evaluation. Notably, these prior works were exclusively validated on extremely limited cohorts (1/3/1 subjects respectively), meaning their evaluations were performed under oversimplified experimental conditions with minimal population diversity. Such constrained validation frameworks cannot sufficiently demonstrate the real-world generalization capabilities of their methodologies.

Autonomous system's efficiency is superior in measurement and comparable in total time: The efficiency of the robotic ultrasound system significantly influences its practical utility. Firstly, as seen in Fig. 5f, the autonomous system's measurement speed substantially exceeds that of sonographers. The most notable reduction in measurement time occurs when the autonomous system decreases the time required by a factor of 64. On average, the autonomous system enhances efficiency by a notable factor of 14. The longer duration taken by sonographers is attributed to the extensive manual operations they undertake, such as selecting high-quality images for measurement, magnifying specific areas, and annotating anatomical landmarks. Secondly, we analyze the total time required for both scanning and measurement by the autonomous system and sonographer. The data reveal that the mean total time between the autonomous system and the sonographer does not significantly differ. Although the autonomous system lags slightly behind the sonographer in terms of total time, the continuous operation of the autonomous system, without the need for breaks, coupled with extended working hours, and its capability for rapid and large-scale replication, predicts a higher efficiency ceiling in clinical scenarios compared to sonographers. Furthermore, the autonomous system demonstrates superior stability in terms of time efficiency.

Subjective and objective comfort assessment of autonomous scanning: Due to the lack of exoskeletal protection in the neck area, along with the crucial function of the carotid arteries in delivering blood to the brain, exerting excessive pressure along the Z-axis (up and down) of the probe during scanning can lead to significant discomfort for the patient, for instance, reduced blood supply to the brain. Upon completion of both autonomous ultrasound scanning and sonographer scanning, participants were invited to rate their comfort on a scale of 0 to 10, where 0 indicates extreme discomfort, and 10

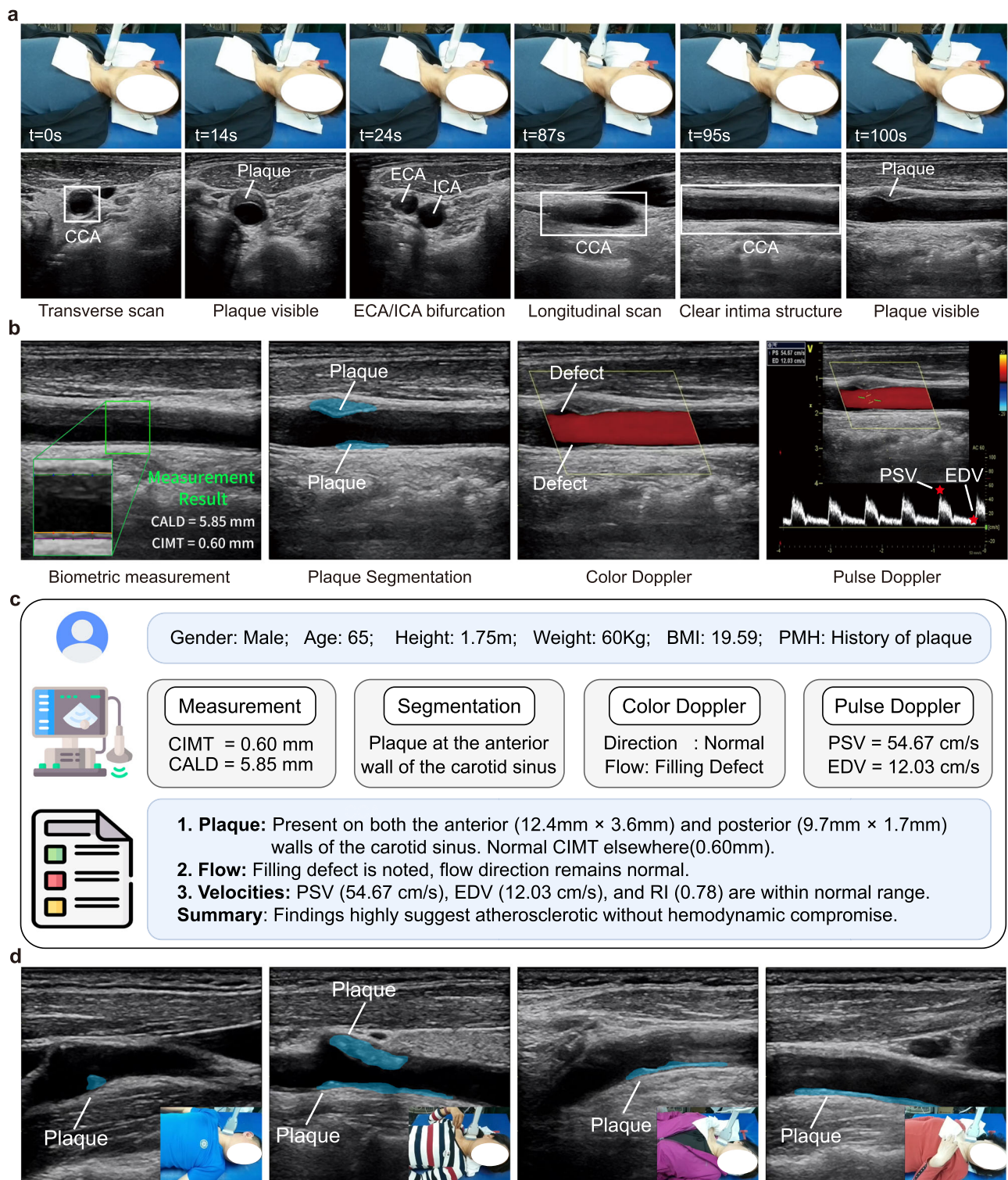
represents utmost comfort. As depicted in Fig. 5g, eight individuals found the comfort level of the autonomous system to be superior to that of the sonographer, whereas seven individuals felt the autonomous system was less comfortable. Overall, the autonomous system received a higher average comfort rating compared to the sonographers. Objectively, using the built-in joint torque sensors and end-effector force estimation application program interface (API) of the Franka robotic arm, we recorded the contact force along the Z-axis. Fig. 5h presents the average force applied by the robotic arm on all participants at various stages. It's observable that the force maintained a comfortable range throughout the entire process. Thus, from both subjective and objective perspectives, the autonomous system showcases satisfactory safety and comfort levels. Technically, we implemented a variable impedance control algorithm to properly balance contact force and control precision, thereby successfully executing tasks while ensuring safety. Please refer to Sec. 5 for more details.

End-to-end autonomous carotid ultrasound process visualization: To offer readers an intuitive grasp of the full process of autonomous examination, we illustrate the scanning, measuring, and segmentation process of an unseen test subject with plaque in Fig. 6. The figure clearly shows the carotid artery remaining consistently visible in the ultrasound images throughout the entire process. In the transverse view, the carotid artery stays centrally positioned, thanks to the model's real-time transducer adjustment. In the longitudinal view, the intimal layer is clearly visualized, demonstrating proper alignment of the ultrasound transducer with the artery's maximal longitudinal cross-section. This optimal orientation provides high-quality images for CALD and CIMT measurements, while also enabling clear visualization of plaques with well-defined borders. Following the acquisition of high-quality ultrasound images, we perform biometric measurements and plaque segmentation to objectively analyze carotid artery anatomy and pathology (Fig. 6b). Using the ultrasound system's intrinsic Color and Pulse Doppler functions, we can further obtain hemodynamic information (Fig. 6b right). The integrated analysis revealed: (1) plaques are present on both the anterior (12.4 mm × 3.6 mm) and posterior (9.7 mm × 1.7 mm) walls of the carotid sinus, with a normal CIMT of 0.60 mm elsewhere, and (2) Doppler ultrasound demonstrating a blood flow filling defect, while flow direction, velocities, and resistive index remained within normal ranges (Fig. 6c). In Fig. 6d, we show four representative plaque segmentation results and report the plaque condition in Supplementary Table 2. We further provide the corresponding Color and Pulse Doppler imaging of these four representative cases in Supplementary Fig. 4. Moreover, three end-to-end examination demonstrations on previously unseen individuals with plaque are available in the Supplementary Videos 1–3.

### Subsystem-level evaluation on expert-annotated datasets

In the context mentioned above, the robotic ultrasound system, comprised of a series of deep neural networks, showcases promising potential for clinical applications. In the subsequent section, for a more comprehensive understanding of each subsystem's performance, we will delve into analyzing their respective functionalities and effectiveness on the test set.

Intelligent action decision based on imitation learning: As introduced earlier in Sec. 5, we collect expert demonstration data and employ an imitation learning strategy to model the decision-making process of sonographers, translating ultrasound images into adjustment actions. Initially, we implement an imitation learning strategy based on deep learning foundational models, namely, ResNet-50<sup>50</sup> and DenseNet-121<sup>51</sup>. Both models yield comparable performance as depicted in Fig. 7a, which illustrates the evaluation of these models on our dataset. Given their equivalent performance, a ResNet-based implementation is employed in all subsequent experiments. Following this, we explore the improvements that deep learning contributes to imitation learning by comparing the ResNet-based implementation with a



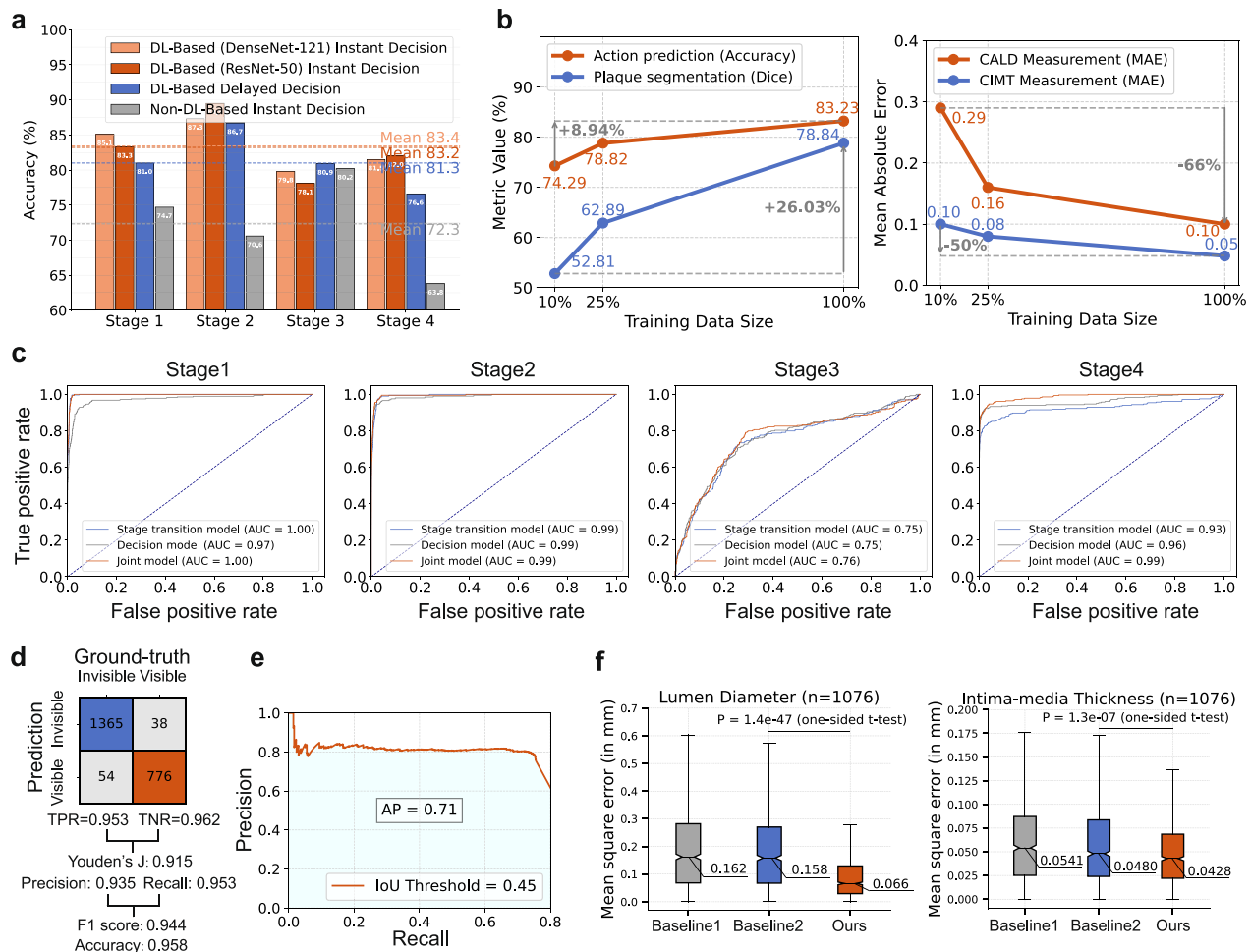
**Fig. 6 | Autonomous end-to-end carotid artery ultrasound examination process visualization.** **a** Scanning process in a 65-year-old patient with plaque, yielding high-quality transverse and longitudinal views with clear plaque visibility. **b** Image analysis of acquired scans, including biometric measurements, plaque segmentation, and color/pulse Doppler. **c** Representative report summarizing scanning and

analysis results. **d** Representative plaque segmentation results of patients. The terms “CCA”, “ICA”, “ECA”, “PMH”, “PSV”, “EDV”, “RI” refer to the common carotid artery, internal carotid artery, external carotid artery, past medical history, peak systolic velocity, end diastolic velocity, and resistive factor, respectively.

non-deep learning method, specifically k-Nearest Neighbors (k-NN). The comparison results in a 10.9% decrease in average accuracy when utilizing the k-NN approach, thereby demonstrating that deep learning significantly enhances the performance of the action decision algorithm. A closer examination reveals that in the two main phases, Stage

2 (transverse scanning of the CCA) and Stage 4 (longitudinal scanning of the CCA), which entail the most complex action decisions, the deep learning approach markedly outperforms k-NN with improvements of 18.9% and 18.2%, respectively. Conversely, in the simpler Stage 3, which encompasses only two action decisions, the deep learning approach





**Fig. 7 | Intelligent action decision and interpretable biometric measurement with deep neural networks.** **a** Comparison of performance between deep learning and non-deep learning method (k-nearest neighbors) in scanning action decision, as well as between instant and delayed decision-making. **b** Performance trends in action decision, biometric measurement, and plaque segmentation as training data scales up. **c** Comparison of the ROC curves for the models of stage transition, action decision, and their combined decision-making in accurately identifying the anatomical structures at the termination positions of each stage. **d** Performance of the model that predicts whether visible arterial wall and intimal structures exist in the

longitudinal section images. **e** Precision-Recall curve of the model in detecting the local region that can be used to measure the arterial structural parameters. **f** Our interpretable biometric measurement solution compared with two other non-interpretable baseline models. The t-test was used to check whether the mean error of our method's results is significantly smaller than that of Baseline 2. The models were evaluated on 1076 annotated images. This boxplot displays standard elements: the box represents the interquartile range (IQR), the central line marks the median, and the whiskers extend to  $1.5 \times \text{IQR}$ .

exhibits slightly weaker performance compared to k-NN. This pattern suggests that deep learning methods are more proficient at navigating in more complex action decision spaces.

One distinctive aspect of ultrasound scanning is the necessity for instantaneous decision-making, which calls for immediate action decisions in response to changes in imaging. To delve deeper into the effects of delays in action decision-making, we envisage a scenario where a decision is repeated twice consecutively. Specifically, the delayed decision-making paradigm executes the action decision  $a_t$  output at time  $t$  twice. Contrarily, our methodology entails making an instant action decision promptly after a single action decision is carried out and there's a subsequent change in the image. In other words, our method outputs and executes an action decision  $a_t$  at time  $t$ . After execution, based on the image at time  $t + 1$ , it further outputs and executes  $a_{t+1}$ . As illustrated in Fig. 7a, there's an absolute reduction of 1.9 percentage points (83.2% vs. 81.3%) in average performance when action decision delays occur. Notably, at the most intricate phase, stage-4, the delayed decision-making strategy significantly impacts the algorithm's efficacy, resulting in an absolute reduction of 5.4 percentage points (82.0% vs. 76.6%). This highlights

the criticality of immediate decision-making in the realm of ultrasound scanning.

In addition to performing a sequence of actions, a pivotal aspect in the successful execution of autonomous ultrasound scanning is pinpointing the termination point for each stage. This primarily entails accurately identifying the anatomical structures depicted in the ultrasound images. To accomplish this identification, we employ two distinct models, namely, the decision model and the stage transition model. The decision model integrates the "stage transition" as a decision action, training it concurrently with the other 12 actions. Meanwhile, the stage transition model serves as a binary classifier, trained separately to ascertain whether transitioning to the subsequent stage is warranted. In Fig. 7c, we showcase the Receiver Operating Characteristic (ROC) curves corresponding to each model, in addition to the outcomes stemming from their collective decision-making process (joint model). Our strategy for joint decision-making is structured such that the progression to the next stage is initiated when either one of the two models opts for stage transition. As illustrated by the figures, the joint model attains superior or at the very least, comparable performance at each stage. To boost the robustness of



identification, we have elected to adopt the joint model as our final implementation.

A core principle of our work is embracing the data scaling law, learning generalizable scanning strategies from extensive expert demonstration data. Thus, we further validate how data scaling improves the model's action decision accuracy. Specifically, we extract 10% and 25% subsets from the complete dataset and trained individual models for each subsample. As shown in Fig. 7b, our results indicate a general upward trend in action prediction accuracy as the dataset expands with more training data. A similar trend has also been observed in biometric measurement and plaque segmentation tasks. This suggests that further data scaling could continue to enhance system performance beyond what is currently presented in our manuscript, reinforcing its potential for reliable deployment.

**Interpretable biometric measurement:** Biometric measurement is another key component, which includes image quality assessment, localized measurement area focusing, and anatomical keypoint detection. Interpretability is a crucial consideration in clinical medicine, as clinicians need to understand and evaluate the correctness of algorithmic outputs. Therefore, we adopt an interpretable approach for predicting CALD and CIMT by regressing points that represent the boundaries of the artery structure. These points can be visualized to allow professionals to easily assess the correctness of the structure (Fig. 8f), and the final biometric measurements are calculated based on these points. Therefore, interpretability is inherently included in this modeling approach. We attempt to compare our approach with non-interpretable methods (Baseline 1/2) in Fig. 7f. Both Baseline 1/2 directly regress the values of CALD and CIMT from the image. The difference lies in their inputs: Baseline 1 takes the complete ultrasound image as input, while Baseline 2 uses a localized region containing clear intimal structures as input (See Supplementary Fig. 5 for visual reference). Compared to our method, the outputs of Baseline 1/2 lack interpretability, making it difficult to assess the correctness of their results in practical applications. As shown in Fig. 7f, it is evident that the interpretable approach outperforms the non-interpretable one. Moreover, the t-test yields a p-value less than 0.001, indicating that the error of the interpretable approach is significantly smaller than that of the non-interpretable one. Additionally, sonographers often zoom in on local regions for more precise measurements during manual procedures. Therefore, we also compare the effects of global versus local regression (Fig. 7f, Baseline 1 vs. Baseline 2). The results indicate that local regression is beneficial for performance.

The performance of the image quality assessment model is presented in Fig. 7d. It can be seen that the model possesses a high recall rate, accurately identifying images with clear intimal structures (Fig. 8d). The performance of the localized measurement area focus model is shown in Fig. 7e. The model demonstrates a high average precision (AP) value, accurately detecting measurable areas (Fig. 8e).

**Precise plaque segmentation:** Carotid artery plaque is one of the most common vascular diseases and a significant risk factor for cardiovascular health. To enable more accurate plaque detection and patient health assessment, we develop an innovative plaque segmentation algorithm. As demonstrated in Fig. 8a, comparative evaluations against state-of-the-art methods<sup>35,54–56</sup> reveal our algorithm's superior performance across all four key metrics (Dice, IoU, Recall, and HD95), while maintaining competitive results in precision. This balanced performance profile suggests our model effectively minimizes false positives while maintaining high detection rates, making it particularly suitable for clinical applications where both accurate plaque identification and reliable negative predictions are equally crucial. Moreover, the vessel region focusing module achieves precise vascular localization, resulting in high AP values (Fig. 8b, c). By incorporating a vessel region focusing mechanism, this algorithm significantly improves segmentation accuracy. Experimental results demonstrate that integrating this module leads to a notable performance improvement, with

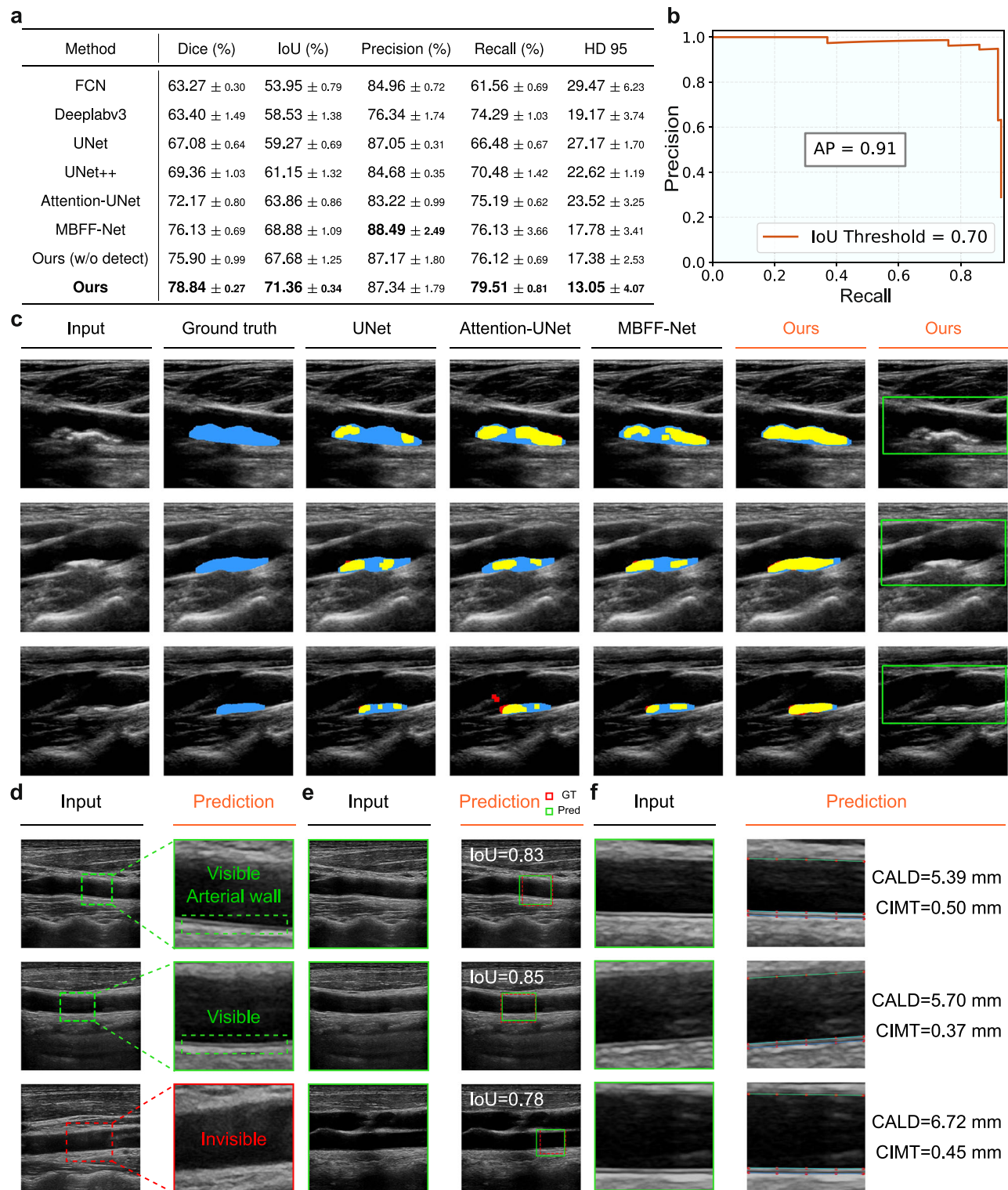
a 2.94% increase in the Dice coefficient (Fig. 8a). Furthermore, our visual comparisons with state-of-the-art methods<sup>35,54,56</sup> demonstrate superior segmentation performance, as shown in Fig. 8c. While competing methods exhibit significant segmentation errors including both under-segmentation (missing plaque regions) and over-segmentation (including non-plaque areas), our approach achieves precise plaque delineation with well-defined boundaries, providing more reliable support for clinical diagnosis and quantitative assessment. Despite achieving promising performance, our segmentation model adheres to the scaling law, as illustrated in Fig. 7b. Notably, the model's performance exhibits no signs of saturation with increasing data volume, suggesting that further improvements can be attained by scaling up the dataset.

## Discussion

In today's healthcare landscape, ultrasound examination stands as one of the most in-demand medical imaging modalities, thanks to its real-time, rapid, and radiation-free features. However, due to its heavy reliance on manual operation, there are issues of non-standardization, inaccuracy, and a marked shortage of qualified professionals<sup>57,58</sup>. These issues are particularly prominent in developing countries<sup>58</sup>, leading to delays in accessing ultrasound examinations and a potential risk of misdiagnosis or missed diagnosis due to low-quality scans. Compared to traditional solutions such as training additional sonographers, developing AI-driven robotic ultrasound systems offers a more promising approach to addressing these global challenges. In recent years, large language models like GPT<sup>29</sup> have demonstrated the scaling law of neural networks<sup>30,31</sup>: as datasets grow, models expand, and learning algorithms advance, new abilities emerge that enable handling complex situations and generalizing to novel samples in “zero-shot” without further training. Inspired by this success, this paper makes the first attempt to demonstrate the great potential of a purely learning-driven autonomous ultrasound robotic system, utilizing large-scale real individual data, more flexible modeling formulations, clinically oriented evaluations, and testing on a large unseen population. Based on the tests conducted on a large, previously unseen population, our robotic system delivers precise biometric results with excellent reproducibility and promising generalizability. Overall, this work provides a proof of concept for developing a learning-based fully autonomous ultrasound robot intended for clinical use.

Future research could probably focus on constructing a unified multimodal perception and decision-making network to further enhance the robotic system's ability to handle complex situations. Multimodal input data can provide a more substantial basis for decision-making, improving the robustness of the system's decisions. Additionally, exploring how to integrate the vast amount of existing ultrasound report data to enhance the capabilities of the robotic system is a potential direction. Multimodal pre-training methods like CLIP<sup>59</sup> have shown that unsupervised training with large volumes of existing paired visual-language data can achieve remarkable generalizability. By utilizing ultrasound reports, the discrepancies in visual appearances of anatomical structures with the same semantics can be aligned through the bridge of language, further enhancing the system's robustness. Lastly, validation on large-scale populations is both time-consuming and labor-intensive, and differences in the populations used for validation across various studies make it difficult to measure technological progress. Thus, researching how to develop a public offline validation tool is worthwhile, as it could potentially accelerate the development of the entire field.

This work is a starting point for intelligent and highly autonomous ultrasound medical robots. In the future, we anticipate ultrasound robots will advance to provide extensive coverage across all organs, be suitable for people of all ages, and seamlessly combine diagnostic and therapeutic functions. We envision a day when ultrasound robots will be capable of performing autonomous full-body scans on patients



diagnostic and treatment capabilities of primary healthcare facilities, thereby enabling more patients to benefit from high-quality medical services. Finally, while the vision outlined above is promising, it requires the collective efforts of researchers in the field of intelligent ultrasound robotics worldwide. We hope our work will ignite their enthusiasm and provide valuable insights for their future work, thereby hastening the fulfillment of this vision.

## Methods

In this section, we introduce our deep learning framework for autonomous ultrasound scanning, biometric measurement, and plaque segmentation, including dataset specifications, training implementation details, as well as the robotic system configuration and the control algorithm. To ensure compliance with ethical standards, all human participant studies were conducted under approval from Tsinghua University's Medical Ethics Committee (THUOI-20230175). We obtained written informed consent from all participants after fully explaining the study procedures, including their understanding that journal publication would include their anonymized data that includes indirect identifiers. Also, the authors affirm that human research participants provided written informed consent, for publication of the images in Figs. 3, 6, and Supplementary Video 1–3.

### Deep learning for autonomous scanning

**Introduction and problem statement.** To reliably collect ultrasound images with high diagnostic value for patients, we draw inspiration from professional sonographers, emulating their policies with our automatic scanning robot. Specifically, the robot aims to capture a transverse view of the upper carotid bifurcation, scanning downward to the lower junction between the common carotid and the subclavian artery. Similarly, the robot is also tasked with scanning the subject in a longitudinal view from the lower junction between the common carotid and the subclavian artery, up to the upper carotid bifurcation. During the scanning process, it is advisable to center on the carotid artery to garner as much information as possible. By adhering to these guidelines, the images obtained can be utilized to make an accurate diagnosis.

**Data collection and annotation.** We recruited 81 volunteers (60 males and 21 females) aged between 18 and 36 years for the collection of expert demonstration data. All data collection sessions were conducted by four qualified sonographers with over 5 years of experience under the supervision of a senior sonographer (15+ years of experience). The data were collected using a General Electric (GE) Vivid E7 ultrasound device equipped with a 9L probe. Sonographers were tasked with performing carotid ultrasound examinations using a series of predetermined keystrokes. The “keystroke actions” refer to a unit-length robot movement in the scanner base coordinate system, totaling 12 (translation and rotation in 3 dimensions) plus 1 (stop action) discrete action space. Since, during data collection, we need to perform one of these actions at each time step, we use a keyboard to conveniently execute these precise actions. Specific keys are programmed and mapped to control a unit movement along one of the 6 degrees of freedom or to stop the scan at a given stage. With every keystroke action, the preceding thirty frames were captured at a rate of 30 FPS before actuating the robot. This process encompassed the recording of both the ultrasound images and their corresponding keystrokes. The sonographers also interacted with the volunteers to ensure their safety and comfort.

To enhance system robustness, multiple initial positions were utilized for imaging the same volunteer. Each volunteer underwent three imaging trajectories following the protocol elaborated in the preceding section. Although volunteers were instructed to remain still during the procedure, minor movements were allowed to account for real-world conditions and to collect data useful for recovery to the optimal view. The comfort level of each participant was continually

monitored, underlining a patient-centric approach. This process thoroughly considered the volunteers' comfort, aiding in the formulation of a comfort-optimized imitation policy. During the process, we also collected expert demonstration data from recovering from poor postures, thereby enhancing the robustness of the model. Finally, we collected 243 trajectory data associated with 247,297 images paired with expert adjustment actions. 76 subjects were grouped into the training set with 231,373 images/action pairs, and 5 subjects were allocated to the test sets with 15,924 images/action pairs. To account for clinical scenarios where multiple actions could be considered valid expert decisions, we implemented a re-annotation protocol under a senior sonographer supervision (15+ years experience). The sonographer was asked to provide up to two additional expert actions for selected validation set images, beyond the single action recorded during actual scanning. This approach recognizes that - similar to how a self-driving car might validly turn either left or right to avoid an obstacle - multiple navigation choices can be clinically appropriate when scanning the carotid artery. While real-world scanning requires choosing a single action, our offline evaluation benefits from considering these clinically equivalent alternatives, leading to more robust performance assessment that better reflects real clinical decision-making.

**Method.** To boost the robustness of the scanning policy, two distinct networks are employed for different operational phases. The initial policy, denoted as  $\pi_{\text{scanning}}(a|o, \theta)$ , where  $a$  represents the discretized robot action, is responsible for outputting the probabilities associated with predefined actions given the observation  $o$ . The predefined action space  $\mathcal{A}$  is illustrated in Fig. 3b. The subsequent policy,  $\pi_{\text{transition}}(a_{\text{transition}}|o, \phi)$ , takes the ultrasound observation  $o$  and outputs  $a_{\text{transition}}$ , which is the probability of transitioning to the subsequent task upon spotting the area of interest. Upon confirmation to transition to the next task, given by  $\max(\pi_{\text{transition}}(a_{\text{transition}}|o, \phi), \mathbf{p}_{\text{transition}}(\pi_{\text{scanning}}(a|o, \theta))) \geq 0.5$ , where  $\mathbf{p}_{\text{transition}}$  picks the probability of transitioning from the action policy, the robot ceases its current operation and moves on to the next stage.

The inputs to these policy networks, as illustrated in Fig. 4b, comprise pre-processed ultrasound images. Initially, the images are cropped to focus exclusively on the ultrasound area and are resized to dimensions of  $3 \cdot 224 \cdot 224$ . It should be noted that the grayscale ultrasound images are expanded into three channels primarily to inherit the models pre-trained on ImageNet, following the practices outlined in references<sup>62,63</sup>. Subsequently, these images are processed using a Convolutional Neural Network (CNN)—specifically, the ResNet-50 architecture<sup>50</sup>, which consists of convolutional blocks with residual connections, followed by average pooling and a fully connected layer. The output dimensions are two for  $\pi_{\text{transition}}$  and thirteen for  $\pi_{\text{scanning}}$ , with a softmax activation function employed to generate action probabilities. It's noteworthy that our method is not exclusive to ResNet and it could be implemented using various networks<sup>51</sup>.

We trained these networks using imitation learning by maximizing the log-probability of the collected ultrasound image  $o$ , along with its corresponding action  $a$ , and checking if it's the transition action  $a_{\text{transition}} = \mathbf{1}_{\text{transition}}(a)$ . Here,  $a \in \mathcal{A}$  and  $a_{\text{transition}} \in [0, 1]$  indicate whether to transition to the next stage or not. The overall objective is expressed as:

$$\max_{\theta, \phi} \mathbb{E}_{o, a \sim \mathcal{D}_{\text{train}}} \left[ \log \pi_{\text{scanning}}(a|o, \theta) + \log \pi_{\text{transition}}(\mathbf{1}_{\text{transition}}(a)|o, \phi) \right]. \quad (1)$$

**Implementation.** Each policy network is trained using a batch size of 256, a learning rate of 0.0001, weight decay of 0.0001, and 10 epochs. The networks are initialized with weights pretrained on ImageNet, and the ultrasound image input is also normalized by ImageNet statistics. The loss function applied to both policies is



cross-entropy, formulated as  $-\sum_{c=1}^M y_{o,a} \log(\pi_{o,a})$ , where  $y$  is the binary indicator verifying whether class label  $a$  is the correct classification for observation  $o$ , and  $\pi$  is the predicted probability that observation  $o$  corresponds to action  $a$ . Both policy networks are trained end-to-end using the Adam optimizer, and employ Distributed Data Parallel (DDP) as implemented by the PyTorch deep learning library.

### Deep learning for biometric measurement

**Introduction and problem statement.** During the diagnostic assessment of carotid artery ultrasound images, sonographers typically concentrate on specific arterial structures. These structures include the boundary between the upper intimal layer and the lumen (intima-up), the boundary between the lower intimal layer (intima-down) and the lumen in carotid ultrasound images, and the boundary between the lower intimal layer and the outer medial layer (media). Specifically, the carotid artery lumen diameter (CALD) is the distance between intima-up and intima-down, while the carotid intima-media thickness (CIMT) is the distance between intima-down and media. These two biometrics are extremely valuable for diagnosing cardiovascular diseases. Our objective is to develop a system capable of automatically calculating CALD and CIMT values based on patients' carotid artery ultrasound images during the autonomous scanning process. Accurate measurement of these biometrics necessitates clear membrane structures within the carotid artery images. However, many ultrasound images in practice are indistinct, posing a challenge to discerning the membrane structures. Consequently, we break down the problem into two stages. The first stage aims to ascertain the presence and location of clear internal membrane structures of the carotid artery in images. If a clear internal membrane structure is detected, the second stage is utilized to calculate CALD and CIMT values. Subsequently, these predicted biometrics assist sonographers in making appropriate diagnostic assessments.

**Data collection and annotation.** Our dataset for biometric measurement is separated into two segments for the training of Stage 1 and Stage 2 models, respectively. Moreover, every image within the dataset is annotated by three expert sonographers.

**Stage 1 Data:** Inspired by the diagnostic practices of sonographers, who typically focus their assessment on particular localized regions in ultrasound images where clear internal membrane structures are discernible, we adjust our data annotation process accordingly. For images showcasing clear internal membrane structures, we delineate regions that exhibit prominent structures of intima-up, intima-down, and media. Following this, we document the coordinates of the top-left and bottom-right points of the bounding box that envelops these features. Conversely, for images devoid of distinct internal membrane structures, the bounding box annotation is omitted. Our Stage 1 dataset encompasses a total of 12,194 images from 76 volunteers. These volunteers are the same individuals used to train the autonomous scanning model. Out of these, 4,957 images display clear internal membrane structures while 7,237 do not. The 76 volunteers are segregated into training and test sets with a 4:1 ratio, ensuring no overlap of individuals across both sets.

**Stage 2 Data:** We notice that prior to measuring the values of CALD and CIMT, sonographers initially determine the positions of intima-up, intima-down, and media in the images by marking points. Motivated by this procedure, we employ the data annotation technique from APRIL<sup>36</sup>, where five equally spaced points are positioned along the  $x$  axis at normalized coordinates [0, 0.25, 0.5, 0.75, 1.0], and their  $y$  coordinates are manually labeled in alignment with the position of the membranes. In other words, each image has a total of 15 key-points labeled. By choosing to predict anatomical boundaries over direct metric values, the interpretability of the model's prediction is augmented. This strategy enables sonographers to visually evaluate the accuracy of the model's predictions during the diagnostic phase, thereby facilitating informed decisions regarding the acceptance or dismissal of the model's results. Our dataset for

training the Stage-2 model consists of 4,957 images. These images are allocated into training and test sets based on individual subjects at a 4:1 ratio, with a guarantee that unique individuals are not included in both sets.

**Method.** Stage 1: A convolution network  $f_{s_1}$  is employed to extract features  $\mathbf{z} \in \mathbb{R}^C$  of a batch of images  $\mathbf{x}_{s_1} \in \mathbb{R}^{3 \times 256 \times 256}$  for multi-task learning. The features are then fed into two distinct heads. The first head ( $h_{cls}$ ) performs binary classification to determine the presence of internal membrane structures. Specifically, for a batch of  $N$  images, it predicts  $\tilde{\mathbf{y}}_s = \{\tilde{y}_{s_i}\}_{i \in \{1, 2, \dots, N\}}$ , each  $\tilde{y}_{s_i}$  indicates a predicted quality score between 0 and 1 of an image in the batch. The second head ( $h_{reg}$ ) predicts  $\tilde{\mathbf{y}}_c = \{\tilde{\mathbf{y}}_{c_i}\}_{i \in \{1, 2, \dots, N\}}$ , each  $\tilde{\mathbf{y}}_{c_i}$  is a 4-dimensional vector, indicating the coordinates of the bounding box's top-left and bottom-right points for an image in the batch.

$$\begin{aligned} \mathbf{z} &= f_{s_1}(\mathbf{x}_{s_1}) \\ \tilde{\mathbf{y}}_s &= h_{cls}(\mathbf{z}) \\ \tilde{\mathbf{y}}_c &= h_{reg}(\mathbf{z}) \end{aligned} \quad (2)$$

During training, the backbone and the two heads are optimized jointly through binary cross entropy (BCE) loss, mean square error (MSE) loss and complete intersection over union (Complete-IoU) loss. Note that the BCE loss is calculated on all images while the MSE loss and the Complete-IoU loss are only calculated on images which are labelled with clear internal membrane structures. The three losses are:

$$\begin{aligned} l_{cls} &= - \sum_{i=1}^N y_{s_i} \cdot \log(\tilde{y}_{s_i}) + (1 - y_{s_i}) \cdot \log(1 - \tilde{y}_{s_i}), \\ l_{reg} &= \sum_{i=1}^N y_{s_i} \cdot \|\mathbf{y}_{c_i} - \tilde{\mathbf{y}}_{c_i}\|^2, \\ l_{IoU} &= \sum_{i=1}^N y_{s_i} \cdot (1 - \text{IoU}(\tilde{\mathbf{y}}_{c_i}, \mathbf{y}_{c_i})). \end{aligned} \quad (3)$$

The overall loss is a weighted sum of these three parts:

$$\mathcal{L}_{s_1} = l_{cls} + \lambda_1 \cdot l_{reg} + \lambda_2 \cdot l_{IoU}. \quad (4)$$

**Stage 2:** Inspired by previous works<sup>64,65</sup>, we formulate the task of keypoint prediction as an offset prediction problem. Instead of directly regressing the coordinates, we first compute the mean positions of each keypoint in the training set as reference points (anchor points), and then predict the offsets to these points, as illustrated in Supplementary Fig. 6. The positional information of these reference points serves as prior knowledge, providing the model with approximate location cues for each keypoint, which has the potential to improve prediction accuracy. For simplicity, we denote the annotated key-points'  $y$  coordinate of the  $i$ -th image in a batch of  $N$  images as  $\mathbf{a}_i$ ,  $\mathbf{b}_i$ ,  $\mathbf{c}_i$ , where  $\mathbf{a}_i = \{a_{ij}\}_{j \in \{1, \dots, 5\}}$ ,  $\mathbf{b}_i = \{b_{ij}\}_{j \in \{1, \dots, 5\}}$ ,  $\mathbf{c}_i = \{c_{ij}\}_{j \in \{1, \dots, 5\}}$ , respectively corresponding to the annotated 5 points on intima of near wall, intima of far wall and media of far wall. Before training, we firstly calculate the mean positions  $\bar{\mathbf{a}}_i = \{\bar{a}_{ij}\}_{j \in \{1, \dots, 5\}}$ ,  $\bar{\mathbf{b}}_i = \{\bar{b}_{ij}\}_{j \in \{1, \dots, 5\}}$ ,  $\bar{\mathbf{c}}_i = \{\bar{c}_{ij}\}_{j \in \{1, \dots, 5\}}$  for each of the point in the training dataset. Then the stage 2 models predicts offsets for each point relative to the corresponding mean positions:

$$\delta_i^a = f_{s_2}^a(\mathbf{x}_{s_2}), \quad \delta_i^b = f_{s_2}^b(\mathbf{x}_{s_2}), \quad \delta_i^c = f_{s_2}^c(\mathbf{x}_{s_2}), \quad (5)$$

where  $\mathbf{x}_{s_2}$  represents the local region detected in Stage 1 that contains clear intima-media structures, with its spatial coordinates determined by  $\tilde{\mathbf{y}}_c$ . Since the  $x$  coordinates are fixed during annotation, only the  $y$  coordinates need to be predicted. The actual  $y$  coordinates are then



computed from the predicted offsets through  $\tilde{a}_{ij} = \bar{a}_{ij} + \delta_{ij}^a$ ,  $\tilde{b}_{ij} = \bar{b}_{ij} + \delta_{ij}^b$ ,  $\tilde{c}_{ij} = \bar{c}_{ij} + \delta_{ij}^c$ . The stage 2 models are optimized using the Mean Squared Error (MSE) loss function:

$$\mathcal{L}_{s_2} = \sum_{i=1}^N \sum_{j=1}^5 \left[ (a_{ij} - \tilde{a}_{ij})^2 + (b_{ij} - \tilde{b}_{ij})^2 + (c_{ij} - \tilde{c}_{ij})^2 \right]. \quad (6)$$

Subsequently, the value of CALD and CIMT can be calculated through the predicted points. Taking CALD as an example, to account for potential non-horizontal positioning of intima-down and intima-up, we fit two lines respectively for these two set of points using the least squares method. Then, the average slope  $k$  of the two lines are obtained and final CALD value can be calculated by  $\tilde{d}_{\text{CALD}} = \frac{1}{5} \sum_{j=1}^5 (a_{ij} - b_{ij}) \cdot \cos(\arctan(k))$ . The CIMT can be calculated similarly.

**Implementation.** For training, distinct models are trained for both Stage 1 and Stage 2. In the first stage of training, the ResNet-50 architecture ( $f_{s_1}$ ) is employed, and the model is trained with a batch size of 256. The initial learning rate is set at 0.001, with a cosine learning rate scheduler applied. The model is trained over 50 epochs, utilizing a weight decay of 0.00001 while conducting 5-fold training. Images are resized to  $256 \cdot 256$  dimensions and normalized using ImageNet statistics. In our experiment, we set  $\lambda_1$  to 1.0 and  $\lambda_2$  to 8.0.

For the second stage of training, the ResNet-50 architecture ( $f_{s_2}$ ) is also employed, and the model is trained with a batch size of 128. The learning rate for this stage is set at 0.0001, following a cosine learning rate scheduler. The model is trained over 100 epochs, with a weight decay set at 0.00001. Images are resized to  $256 \cdot 256$  dimensions and normalized using a mean of 0.193 and a standard deviation of 0.224.

During the autonomous scanning procedure, the system captures real-time ultrasound images of the carotid artery. The first stage model is deployed to evaluate the presence and location of internal membrane structures within the acquired image. Upon identification of such structures, the image is cropped using the predicted bounding box coordinates. This cropped image is then supplied to the second stage model for key-point prediction and subsequent biometric calculations.

## Deep learning for plaque segmentation

**Introduction and problem statement.** Carotid atherosclerotic plaques are a major risk factor for ischemic stroke and other related diseases. Accurate segmentation of plaques in medical images is of great importance for the diagnosis of carotid artery conditions. However, the inherent low contrast of ultrasound images, the presence of noise, and the high heterogeneity of plaque morphology pose significant challenges to precise segmentation. Clinical observations show that plaques often form between the intima and media layers of the carotid artery wall. When analyzing ultrasound images, clinicians typically locate the vessel wall first, then identify abnormal protrusions along its boundary. Inspired by this diagnostic process, we propose a dual-branch, multi-scale segmentation network with explicit vascular spatial priors, as illustrated in Fig. 4d. The framework introduces a vessel region localization stage, where a detection model is trained to capture spatial cues of the vessel region. These spatial priors are then embedded into the segmentation model to guide learning, helping the model focus on plausible plaque areas and suppress irrelevant background noise.

**Data collection and annotation.** The dataset used for plaque segmentation experiments, adopted from the MBFF-Net<sup>35</sup> study, consists of 430 carotid ultrasound images collected from different patients. These images were acquired using two ultrasound systems: the Philips IU22 with an L9-3 probe and the GE Logiq E9 with a 9L probe, both operating at a center frequency of 9 MHz. Each image is accompanied by a manually annotated plaque mask, and we further labeled bounding boxes to delineate vascular regions. Before being input into

the model, all images are resized to  $256 \cdot 256$  pixels. Among the dataset, 330 images are used for training and 100 images for testing.

**Method.** Given an input ultrasound image  $\mathbf{I} \in \mathbb{R}^{3 \cdot H \cdot W}$ , we use ResNeXt<sup>66</sup> to extract multi-scale hierarchical features  $\{\mathbf{d}_i\}_{i=1}^4$ ,  $\mathbf{d}_i \in \mathbb{R}^{C_i \cdot H_i \cdot W_i}$ . The hierarchical features are then processed through two parallel branches: a vessel region-focused branch and a multi-scale feature integration branch. Please refer to Supplementary Fig. 7 for more intuitive understanding of the network architecture.

**Branch 1:** Since plaques typically grow along the arterial wall, we incorporate vascular wall location priors into the model to reduce interference from irrelevant regions. Specifically, we trained a Faster R-CNN<sup>41</sup> model to detect the vascular region in the input image, producing bounding box coordinates  $(x_{\min}, y_{\min}, x_{\max}, y_{\max})$ . We then generate a masked input image  $\mathbf{Z}$  by setting pixels outside the bounding box to zero:

$$\mathbf{Z}_{x,y} = \begin{cases} \mathbf{I}_{x,y} & \text{if } x_{\min} \leq x \leq x_{\max} \text{ and } y_{\min} \leq y \leq y_{\max}, \\ 0 & \text{else.} \end{cases} \quad (7)$$

Then,  $\mathbf{Z}$  is encoded by convolutional layers and fed into the spatial feature transform<sup>67</sup> module, which undergoes affine transformation to generate the modulation parameters  $\gamma_i \in \mathbb{R}^1$ ,  $\beta_i \in \mathbb{R}^1$ , and finally outputs the modulation features  $\{\mathbf{e}_i^1\}_{i=1}^4$ :

$$\mathbf{e}_i^1 = (\gamma_i + 1)\mathbf{d}_i + \beta_i. \quad (8)$$

**Branch 2:** Meanwhile, the multi-scale features  $\{\mathbf{d}_i\}_{i=1}^4$  are concatenated along the channel dimension and then passed through a fusion layer to generate the global contextual feature  $\mathbf{g}$ . Subsequently,  $\mathbf{g}$  is concatenated with each multi-scale feature  $\mathbf{d}_i$ , and the combined features are processed through convolutional and activation layers to produce the second branch output features  $\{\mathbf{e}_i^2\}_{i=1}^4$ .

For each hierarchical level  $i \in \{1, 2, 3, 4\}$ , the output features  $\mathbf{e}_i^1$  and  $\mathbf{e}_i^2$  from both branches are fused. The combined features are then processed through multi-layer convolutions and upsampling operations to generate the final prediction result  $\mathbf{p}_i$ . The final generated output  $\mathbf{p}_i$  maintains the same spatial size as the original input image, i.e.,  $256 \cdot 256$ . Subsequently, the binary cross-entropy loss  $\mathcal{L}_i$  is computed between each hierarchical prediction  $\mathbf{p}_i$  and the ground-truth segmentation mask  $\mathbf{S}_{\text{gt}}$ :

$$\mathcal{L}_i = -\frac{1}{N} \sum_{x,y} \left[ \mathbf{S}_{\text{gt}}(x,y) \log \sigma(\mathbf{p}_i(x,y)) + (1 - \mathbf{S}_{\text{gt}}(x,y)) \log(1 - \sigma(\mathbf{p}_i(x,y))) \right], \quad (9)$$

where the  $\sigma(\cdot)$  represents the sigmoid activation function. Finally, the total loss is obtained by summing the individual losses across all hierarchical levels:  $\mathcal{L}_{\text{total}} = \sum_{i=1}^4 \mathcal{L}_i$ . During model inference, the average of predictions  $\{\mathbf{p}_i\}_{i=1}^4$  from all hierarchical levels is computed as the final prediction result:

$$\mathbf{S}_{\text{final}} = \sigma \left( \frac{1}{4} \sum_{i=1}^4 \mathbf{p}_i \right) \quad (10)$$

**Implementation.** The training process of the network comprises two stages. In the first stage, a pretrained Faster R-CNN model with a ResNet-50-FPN backbone is fine-tuned for the vessel wall detection task. In the second stage, both the original images and their corresponding vessel wall detection results are used as inputs to train the segmentation model. Identical training parameters are applied to both stages: a batch size of 2, initial learning rate of 0.005, SGD optimizer with momentum 0.9, weight decay of 0.0005, and a total of 100 training epochs. All experiments are conducted on a single RTX3090 GPU.

## Robotic system configuration

The robotic ultrasound system consists of a 7-degree-of-freedom collaborative robotic arm (Franka Emika Panda) with a General Electric (GE) Vivid E7 ultrasound device equipped with a 9L probe. The probe is rigidly attached to the robotic arm's end effector, enabling precise control of the probe's position and orientation. Before scanning, we apply an adequate amount of ultrasound gel evenly on the probe surface to ensure optimal acoustic coupling between the probe and the subject's skin, which is essential for high-quality ultrasound imaging. The ultrasound imaging parameters, including gain and dynamic range, are preset to 6 and 72, respectively, providing clear and consistent image quality. The method of guiding the probe from its initial position to contact the human neck using an external depth camera has been well-established in our previous works<sup>68</sup>. Readers may refer to existing literature<sup>68</sup> for technical details. During scanning, since GE does not provide direct software access to the imaging data, we use a high-performance video capture card (Acasis, Shenzhen, China) to record the ultrasound monitor's display. The captured video stream is then processed by extracting the region containing the ultrasound image, which is subsequently fed into the neural network for analysis.

## Robot control algorithm

Cartesian impedance control<sup>69</sup> is employed during ultrasound scanning to execute motion commands. This control strategy ensures compliant and stable interaction between the robotic arm and the subject's body, enabling precise probe positioning while maintaining safe contact forces. Specifically, the controller we utilize is a streamlined version of the traditional Cartesian impedance control:

$$\tau = J^T(-K\tilde{x} - D\dot{\tilde{x}}) + (I - J^T J^{+T})(-K_n\dot{q} - D_n\ddot{q}) + C\dot{q} + g, \quad (11)$$

where  $\tilde{x} = x - x_d \in \mathbb{R}^6$  represents the pose (position and orientation) error of the probe in Cartesian space, the subscript  $d$  for desired,  $\dot{\tilde{x}}$  signifies the velocity of pose. The vectors  $q, \dot{q}, \ddot{q} \in \mathbb{R}^7$  correspond to the joint angle, joint velocity and joint error. The Jacobian matrix  $J \in \mathbb{R}^{6 \times 7}$  maps from joint space to Cartesian space, and the superscript  $+$  indicates the pseudo-inverse. The stiffness matrices  $K$  and  $K_n$  correspond to the Cartesian space and null space, respectively, while the damping matrices  $D$  and  $D_n$  are set to ensure critical damping. The terms  $C\dot{q}$  and  $g$  account for Coriolis and gravitational forces, respectively.

When we substitute (11) into the dynamics model of the robot

$$M\ddot{q} + C\dot{q} + g = \tau + \tau_{\text{ext}}, \quad (12)$$

where  $M \in \mathbb{R}^{7 \times 7}$  denotes the mass matrix, and  $\tau, \tau_{\text{ext}} \in \mathbb{R}^7$  represent the control torque and the external torque, respectively, we have the close-loop dynamics

$$M\ddot{q} + J^T(K\tilde{x} + D\dot{\tilde{x}}) + (I - J^T J^{+T})(K_n\dot{q} + D_n\ddot{q}) = \tau_{\text{ext}}. \quad (13)$$

To derive the dynamics of the probe, we left-multiply (13) by the Jacobian  $J$ . Utilizing the null-space projection property  $J(I - J^T J^{+T}) = 0$ , we obtain:

$$JM\ddot{q} + JJ^T(K\tilde{x} + D\dot{\tilde{x}}) = J\tau_{\text{ext}}. \quad (14)$$

During quasi-static motion, including the equilibrium state during probe-neck contact, we have  $\ddot{q} = 0, \dot{\tilde{x}} = 0$ . Therefore,

$$K\tilde{x} = (JJ^T)^{-1}J\tau_{\text{ext}} = J^{+T}\tau_{\text{ext}} = F_{\text{ext}}, \quad (15)$$

where  $F_{\text{ext}}$  is the contact force between the probe and the patient's neck. Equation (15) reveals that the contact force increases

proportionally and thus unboundedly with the pose error, posing a safety risk in practical implementation.

To address this safety concern, we introduce a modification where the stiffness matrix  $K$  becomes error-dependent. We established a safe threshold for the contact force, denoted by  $F_{\text{ext}} = [\bar{f}_{\text{ext}1}, \bar{f}_{\text{ext}2}, \dots, \bar{f}_{\text{ext}6}]^T$ . If the contact force exceeds this threshold, the stiffness matrix  $K$  is adjusted to maintain the contact force within a safe range, as defined by:

$$k = \begin{cases} k_{\text{normal}}, & \text{if } f_{\text{ext}} < \bar{f}_{\text{ext}}, \\ \bar{f}_{\text{ext}}/\tilde{x}, & \text{if } f_{\text{ext}} \geq \bar{f}_{\text{ext}}, \end{cases} \quad (16)$$

where  $k$  and  $\tilde{x}$  represent the corresponding elements of the stiffness matrix and the pose error vector, respectively.

Additionally, regarding safe human-robot interaction scenarios—such as a human attempting to push away a robotic arm or accidental collisions between the robot and other humans—our team has conducted detailed studies in the prior work<sup>68</sup>. This work proposed a safety interaction framework to address these cases. As this falls outside the scope of the current paper, readers may refer to<sup>68</sup> for further details.

## Statistics and reproducibility

No statistical method was used to predetermine sample size. To validate the performance of our robotic system, we recruited 41 volunteers with diverse demographic characteristics. In this test cohort, the oldest participant was 70 years old, including 7 subjects over 60 (6 of whom exhibited plaques), 7 aged between 45–60, and the remaining under 45. The volunteers exhibited a broad spectrum of physiques: heights ranged from 1.55 to 1.90 m (mean  $\pm$  SD =  $1.72 \pm 0.09$  m), weights varied between 46.0–100.0 kg ( $65.0 \pm 12.1$  kg), and BMI spanned 16.5–30.8 ( $22.1 \pm 3.2$ ), with 12 subjects having BMI <20 and 13 subjects BMI  $\geq 24$ . This population diversity ensures robust evaluation of the system's real-world performance and enhances the reproducibility of our findings.

## Reporting summary

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

## Data availability

To ensure compliance with participant agreements and prevent commercial misuse, all datasets are available under controlled access. All data requests must include: (1) institutional affiliation details, (2) a research purpose statement, and (3) a signed [data use agreement](#). An independent review panel will consider and approve requests for verified academic research purposes. For data inquiries, please contact the lead author (Haojun Jiang; [jhj20@mails.tsinghua.edu.cn](mailto:jhj20@mails.tsinghua.edu.cn)). All such requests will be processed within two weeks. Source data are provided with this paper.

## Code availability

The code for this project is available on GitHub repository: <https://github.com/LeapLabTHU/UltraBot>.

## References

- Salomon, L. J. et al. Practice guidelines for performance of the routine mid-trimester fetal ultrasound scan. *Ultrasound Obstet. Gynecol.* **37**, 116–126 (2011).
- Namburete, A. I. et al. Normative spatiotemporal fetal brain maturation with satisfactory development at 2 years. *Nature* 1–9 (2023).
- Ulloa Cerna, A. E. et al. Deep-learning-assisted analysis of echocardiographic videos improves predictions of all-cause mortality. *Nat. Biomed. Eng.* **5**, 546–554 (2021).

4. Lin, M. et al. A fully integrated wearable ultrasound system to monitor deep tissues in moving subjects. *Nature Biotechnology* **1**–10 (2023).
5. Stein, J. H. et al. Use of carotid ultrasound to identify subclinical vascular disease and evaluate cardiovascular disease risk: a consensus statement from the american society of echocardiography carotid intima-media thickness task force endorsed by the society for vascular medicine. *J. Am. Soc. Echocardiogr.* **21**, 93–111 (2008).
6. Hu, H. et al. A wearable cardiac ultrasound imager. *Nature* **613**, 667–675 (2023).
7. Ferraioli, G. & Monteiro, L. B. S. Ultrasound-based techniques for the diagnosis of liver steatosis. *World J. Gastroenterol.* **25**, 6053 (2019).
8. Ferraioli, G. et al. Liver ultrasound elastography: an update to the world federation for ultrasound in medicine and biology guidelines and recommendations. *Ultrasound Med. Biol.* **44**, 2419–2440 (2018).
9. Wang, C. et al. Monitoring of the central blood pressure waveform via a conformal ultrasonic device. *Nat. Biomed. Eng.* **2**, 687–695 (2018).
10. Thomson, N. Sonographer workforce survey analysis. *Society of Radiographers* (2014).
11. Beales, L., Wolstenhulme, S., Evans, J., West, R. & Scott, D. Reproducibility of ultrasound measurement of the abdominal aorta. *J. Br. Surg.* **98**, 1517–1525 (2011).
12. Joakimsen, O., Bønaa, K. H. & Stensland-Bugge, E. Reproducibility of ultrasound assessment of carotid plaque occurrence, thickness, and morphology: the tromsø study. *Stroke* **28**, 2201–2207 (1997).
13. Parker, P. & Harrison, G. Educating the future sonographic workforce: Membership survey report from the british medical ultrasound society. *Ultrasound* **23**, 231–241 (2015).
14. Committee, M. A. Skilled shortage sensible: full review of the recommended shortage occupation lists for the uk and scotland, a sunset clause and the creative occupations. *Migration Advisory Committee* (2013).
15. Buonsenso, D., Pata, D. & Chiaretti, A. Covid-19 outbreak: less stethoscope, more ultrasound. *Lancet Respiratory Med.* **8**, e27 (2020).
16. Gargani, L. et al. Why, when, and how to use lung ultrasound during the covid-19 pandemic: enthusiasm and caution. *Eur. Heart J.-Cardiovascular Imaging* **21**, 941–948 (2020).
17. Tahmasebpour, H. R., Buckley, A. R., Cooperberg, P. L. & Fix, C. H. Sonographic examination of the carotid arteries. *Radiographics* **25**, 1561–1575 (2005).
18. Wendelhag, I., Gustavsson, T., Suurkula, M., Berglund, G. & Wikstrand, J. Ultrasound measurement of wall thickness in the carotid artery: fundamental principles and description of a computerized analysing system. *Clin. Physiol.* **11**, 565–577 (1991).
19. Oates, C. et al. Joint recommendations for reporting carotid ultrasound investigations in the united kingdom. *Eur. J. Vasc. Endovasc. Surg.* **37**, 251–261 (2009).
20. Song, P. et al. Global and regional prevalence, burden, and risk factors for carotid atherosclerosis: a systematic review, meta-analysis, and modelling study. *Lancet Glob. Health* **8**, e721–e729 (2020).
21. O’Leary, D. H. & Bots, M. L. Imaging of atherosclerosis: carotid intima-media thickness. *Eur. heart J.* **31**, 1682–1689 (2010).
22. Roth, G. A. et al. Global, regional, and national burden of cardiovascular diseases for 10 causes, 1990 to 2015. *J. Am. Coll. Cardiol.* **70**, 1–25 (2017).
23. Huang, D., Bi, Y., Navab, N. & Jiang, Z. Motion magnification in robotic sonography: enabling pulsation-aware artery segmentation. *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* 6565–6570 (2023).
24. Huang, Y. et al. Towards fully autonomous ultrasound scanning robot with imitation learning based on clinical protocols. *IEEE Robot. Autom. Lett.* **6**, 3671–3678 (2021).
25. Huang, Q., Gao, B. & Wang, M. Robot-assisted autonomous ultrasound imaging for carotid artery. *IEEE Trans. Instrum. Meas.* **73**, 1–9 (2024).
26. Bi, Y. et al. Vesnet-rl: Simulation-based reinforcement learning for real-world us probe navigation. *IEEE Robot. Autom. Lett.* **7**, 6638–6645 (2022).
27. Su, K. et al. A fully autonomous robotic ultrasound system for thyroid scanning. *Nat. Commun.* **15** 1, 4004 (2024).
28. Ning, G., Zhang, X. & Liao, H. Autonomic robotic ultrasound imaging system based on reinforcement learning. *IEEE Trans. Biomed. Eng.* **68**, 2787–2797 (2021).
29. Brown, T. et al. Language models are few-shot learners. *Adv. neural Inf. Process. Syst.* **33**, 1877–1901 (2020).
30. Kaplan, J. et al. Scaling laws for neural language models. *arXiv preprint arXiv:2001.08361* (2020).
31. Zhai, X., Kolesnikov, A., Houtsby, N. & Beyer, L. Scaling vision transformers. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* 12104–12113 (2022).
32. Ross, S., Gordon, G. & Bagnell, D. A reduction of imitation learning and structured prediction to no-regret online learning. *Proceedings of the fourteenth international conference on artificial intelligence and statistics* 627–635 (2011).
33. Triantafyllidis, E., Acero, F., Liu, Z. & Li, Z. Hybrid hierarchical learning for solving complex sequential tasks using the robotic manipulation network roman. *Nature Machine Intelligence* 1–15 (2023).
34. Hussein, A., Gaber, M. M., Elyan, E. & Jayne, C. Imitation learning: A survey of learning methods. *ACM Comput. Surv. (CSUR)* **50**, 1–35 (2017).
35. Mi, S., Bao, Q., Wei, Z., Xu, F. & Yang, W. Mbff-net: Multi-branch feature fusion network for carotid plaque segmentation in ultrasound. *Medical image computing and computer-assisted intervention* 313–322 (2021).
36. Lian, S., Luo, Z., Feng, C., Li, S. & Li, S. April: Anatomical prior-guided reinforcement learning for accurate carotid lumen diameter and intima-media thickness measurement. *Med. Image Anal.* **71**, 102040 (2021).
37. Johri, A. M. et al. Recommendations for the assessment of carotid arterial plaque by ultrasound for the characterization of atherosclerosis and evaluation of cardiovascular risk: from the american society of echocardiography. *J. Am. Soc. Echocardiogr.* **33**, 917–933 (2020).
38. Lee, W. General principles of carotid doppler ultrasonography. *Ultrasonography* **33**, 11 (2014).
39. Vellido, A. The importance of interpretability and visualization in machine learning for applications in medicine and health care. *Neural Comput. Appl.* **32**, 18069–18083 (2020).
40. Stiglic, G. et al. Interpretability of machine learning-based prediction models in healthcare. *Wiley Interdiscip. Rev.: Data Min. Knowl. Discov.* **10**, e1379 (2020).
41. Ren, S., He, K., Girshick, R. & Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems* **28** (2015).
42. Abolmaesumi, P., Salcudean, S. E., Zhu, W.-H., Sirouspour, M. R. & DiMaio, S. P. Image-guided control of a robot for medical ultrasound. *IEEE Trans. Robot. Autom.* **18**, 11–23 (2002).
43. Fang, T.-Y., Zhang, H. K., Finocchi, R., Taylor, R. H. & Boctor, E. M. Force-assisted ultrasound imaging system through dual force sensing and admittance robot control. *Int. J. computer Assist. Radiol. Surg.* **12**, 983–991 (2017).
44. Welleweerd, M. K., de Groot, A. G., de Looijer, S., Siepel, F. J. & Stramigioli, S. Automated robotic breast ultrasound acquisition using ultrasound feedback. *2020 IEEE international conference on robotics and automation (ICRA)* 9946–9952 (2020).



45. Li, K. et al. Autonomous navigation of an ultrasound probe towards standard scan planes with deep reinforcement learning. *2021 IEEE International Conference on Robotics and Automation (ICRA)* 8302–8308 (2021).
46. Zhan, J., Cartucho, J. & Giannarou, S. Autonomous tissue scanning under free-form motion for intraoperative tissue characterisation. *2020 IEEE international conference on robotics and automation (ICRA)* 11147–11154 (2020).
47. Peters, S. et al. Manual or semi-automated edge detection of the maximal far wall common carotid intima-media thickness: a direct comparison. *J. Intern. Med.* **271**, 247–256 (2012).
48. Freire, C. M. V. et al. Comparison between automated and manual measurements of carotid intima-media thickness in clinical practice. *Vascular health and risk management* 811–817 (2009).
49. Schuirmann, D. J. A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *J. Pharmacokinetics Biopharmaceutics* **15**, 657–680 (1987).
50. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition* 770–778 (2016).
51. Huang, G., Liu, Z., Pleiss, G., Van Der Maaten, L. & Weinberger, K. Q. Convolutional networks with dense connectivity. *IEEE Trans. pattern Anal. Mach. Intell.* **44**, 8704–8716 (2019).
52. Long, J., Shelhamer, E. & Darrell, T. Fully convolutional networks for semantic segmentation. *Proceedings of the IEEE conference on computer vision and pattern recognition* 3431–3440 (2015).
53. Chen, L.-C., Papandreou, G., Schroff, F. & Adam, H. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587* (2017).
54. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. *Medical image computing and computer-assisted intervention* 234–241 (2015).
55. Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N. & Liang, J. Unet++: A nested u-net architecture for medical image segmentation. *Deep learning in medical image analysis and multimodal learning for clinical decision support* 3–11 (2018).
56. Oktay, O. et al. Attention u-net: Learning where to look for the pancreas. *Medical Imaging with Deep Learning* (2018).
57. Won, D. et al. Sound the alarm: The sonographer shortage is echoing across healthcare. *Journal of Ultrasound in Medicine* (2024).
58. Shah, S. et al. Perceived barriers in the use of ultrasound in developing countries. *Crit. ultrasound J.* **7**, 1–5 (2015).
59. Radford, A. et al. Learning transferable visual models from natural language supervision. *International conference on machine learning* 8748–8763 (2021).
60. Jennings, P., Coral, A., Donald, J., Rode, J. & Lees, W. Ultrasound-guided core biopsy. *Lancet* **333**, 1369–1371 (1989).
61. Zhang, M. et al. Ultrasound-guided radiofrequency ablation versus surgery for low-risk papillary thyroid microcarcinoma: results of over 5 years' follow-up. *Thyroid* **30**, 408–417 (2020).
62. Christiansen, F. et al. International multicenter validation of ai-driven ultrasound detection of ovarian cancer. *Nature Medicine* 1–8 (2025).
63. Qian, X. et al. A multimodal machine learning model for the stratification of breast cancer risk. *Nature Biomedical Engineering* 1–15 (2024).
64. Law, H. & Deng, J. Cornernet: Detecting objects as paired keypoints. *Proceedings of the European conference on computer vision* 734–750 (2018).
65. Duan, K. et al. Centernet: Keypoint triplets for object detection. *Proceedings of the IEEE/CVF international conference on computer vision* 6569–6578 (2019).
66. Xie, S., Girshick, R., Dollár, P., Tu, Z. & He, K. Aggregated residual transformations for deep neural networks. *Proceedings of the IEEE conference on computer vision and pattern recognition* 1492–1500 (2017).
67. Wang, X., Yu, K., Dong, C. & Change Loy, C. Recovering realistic texture in image super-resolution by deep spatial feature transform. *Proceedings of the IEEE conference on computer vision and pattern recognition* 606–615 (2018).
68. Yan, X. et al. A unified interaction control framework for safe robotic ultrasound scanning with human-intention-aware compliance. *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* 14004–14011 (2024).
69. Albu-Schaffer, A. & Hirzinger, G. Cartesian impedance control techniques for torque controlled light-weight robots. *2002 IEEE Int. Conf. Robot. Autom. (ICR A)* **1**, 657–663 (2002).

## Acknowledgements

Gao Huang is supported by the National Key R&D Program of China under Grant 2024YFB4708200 and the Scientific Research Innovation Capability Support Project for Young Faculty of MoE of China under Grant ZYGXQNJSKYCXNLZCXM-I20. Qian Yang is supported by the Key Research of National Social Science Foundation of China under Grant 2024-SKJJ-B-047 and Comprehensive Research on Air Force Equipment Foundation under Grant KJ2023COKYD19. Xiang Li is supported by the National Natural Science Foundation of China under Grant 62461160307. Wenming Yang is supported by the Special Foundations for the Development of Strategic Emerging Industries of Shenzhen under Grant No.KJZD20231023094700001. We sincerely thank Professor Jianwen Luo and Mr. Rui Wang for providing the EQTouch ultrasound machine for our experiments. We are also grateful to Professor Guangyu Wang and Dr. Siqi Zhang for their valuable suggestions during the rebuttal phase.

## Author contributions

H.J. led the project; H.J., G.H., and A.Z. contributed to the conception of the study; Q.Y. and K.H. provided critical medical expertise and conceptual guidance throughout the project; H.J., A.Z., Q.Y., J.W., H.W., N.J., and S.L. contributed to the expert scanning demonstration data collection; H.J., A.Z., and G.H. designed the scanning action decision-making algorithm and the on-human experimental protocol; H.J., Q.Y., A.Z., T.W., X.Y., N.J., L.R., S.C., and G.Y. participated in executing the human trials; H.J. led the data analysis of the human trial results, with participation from A.Z.; A.Z. and H.J. performed off-line evaluation of scanning algorithm; H.J., J.W., and P.L. contributed to the collection of the biometric measurement dataset; H.J., J.W., and G.H. designed the biometric measurement algorithm; H.J., J.W., and T.W. contributed to the evaluation of biometric measurement algorithm; W.Y., G.H., H.J., and T.W. contributed to the collection of the plaque segmentation dataset; H.J., T.W., and G.H. designed the plaque segmentation algorithm; H.J. and T.W. contributed to the evaluation of plaque segmentation algorithm; X.Y., S.L., G.W., and X.L. contributed to the robotic control algorithm; H.J., Y.W., A.Z., T.W., X.Y., and J.W. wrote the manuscript; G.H., K.H., Y.W., S.S., and X.L. helped perform the analysis and write the manuscript with constructive discussions; T.W., H.J., and A.Z. contributed to the demonstration video production process; N.J. and Y.Y. contributed to pilot experiments; G.H. and K.H. supervised the work.

## Competing interests

The authors declare no competing interests.



## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-62865-w>.

**Correspondence** and requests for materials should be addressed to Kunlun He or Gao Huang.

**Peer review information** *Nature Communications* thanks Diego Dal-l'Alba, Floris Ernst, and Guillaume Goudot for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

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

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025, corrected publication 2025