scientific reports

Check for updates

OPEN

Transforming obstetric ultrasound into data science using eye tracking, voice recording, transducer motion and ultrasound video

Lior Drukker[1,3], Harshita Sharma[2,3], Richard Droste[2], Mohammad Alsharid[2], Pierre Chatelain[2], J. Alison Noble[2] & Aris T. Papageorghiou[1]

Ultrasound is the primary modality for obstetric imaging and is highly sonographer dependent. Long training period, insufficient recruitment and poor retention of sonographers are among the global challenges in the expansion of ultrasound use. For the past several decades, technical advancements in clinical obstetric ultrasound scanning have largely concerned improving image quality and processing speed. By contrast, sonographers have been acquiring ultrasound images in a similar fashion for several decades. The PULSE (Perception Ultrasound by Learning Sonographer Experience) project is an interdisciplinary multi-modal imaging study aiming to offer clinical sonography insights and transform the process of obstetric ultrasound acquisition and image analysis by applying deep learning to large-scale multi-modal clinical data. A key novelty of the study is that we record full-length ultrasound video with concurrent tracking of the sonographer's eyes, voice and the transducer while performing routine obstetric scans on pregnant women. We provide a detailed description of the novel acquisition system and illustrate how our data can be used to describe clinical ultrasound. Being able to measure different sonographer actions or model tasks will lead to a better understanding of several topics including how to effectively train new sonographers, monitor the learning progress, and enhance the scanning workflow of experts.

There are few population screening programs that rely on imaging as the primary screening modality—mammography, aortic aneurysm screening and screening during pregnancy are three[1]. Of these, worldwide, obstetric ultrasound is by far the most used. Ultrasound is a relatively low-cost medical imaging modality, compared to X-ray, CT and MRI, is convenient and painless, does not use ionizing radiation, yields immediate results, and is widely considered to be safe[2]. In the past decade and by virtue of advancement in the understanding of fetal functional anatomy, standardization of scanning guidelines, clinical diagnosis, and technical developments of image acquisition equipment, the quality of obstetric ultrasound has improved. In particular, advances in ultrasound physics, materials science, electronics, and computational power have collectively resulted in improvements in image quality and capabilities: higher spatial and temporal resolution as well as increased signal-to-noise ratio, minimization of artifacts, and three-dimensional (3D) rendering[3-9].

Unlike these achievements, the process of ultrasound scanning, the technique of finding standard anatomical planes of the fetus to allow their diagnostic examination, remains relatively unchanged. Routine obstetric ultrasound scans are performed by a sonographer sitting or standing next to a pregnant woman lying down, manipulating a transducer and adjusting the machine settings, following a defined protocol, in order to acquire a series of two-dimensional standard imaging planes, observed on the screen of the ultrasound machine. This process can be thought of as the "sonographer loop": the sonographer moves the transducer looking for a standard plane, receives real-time visual feedback of the video on the display screen and constantly moves their hand to adjust the transducer position, which changes the displayed video (Fig. 1).

[1]Nuffield Department of Women's and Reproductive Health, University of Oxford, Oxford, UK. [2]Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Oxford, UK. [3]These authors contributed equally: Lior Drukker and Harshita Sharma. [4]These authors are joint senior authors: J. Alison Noble and Aris T. Papageorghiou. ✉email: alison.noble@eng.ox.ac.uk
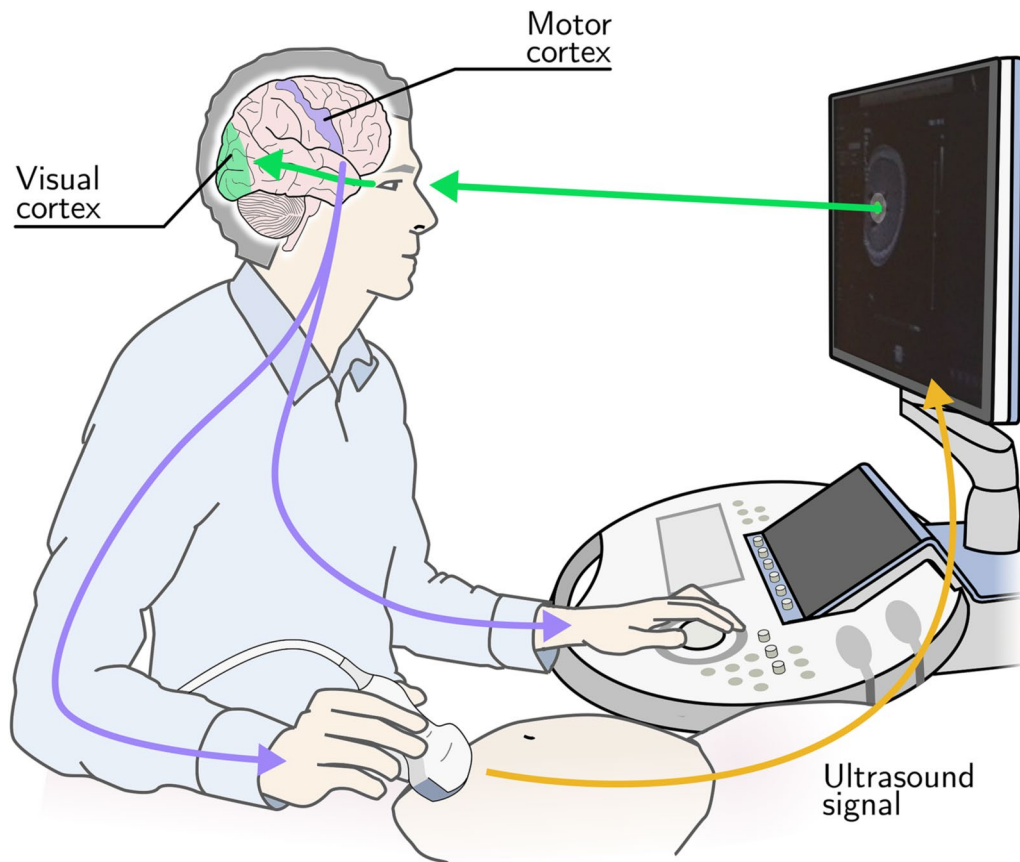
Figure 1. "Sonographer loop" is the process ultrasound sonographers undertake while acquiring standard planes: the sonographer gazes at the ultrasound screen, manipulates the hand to adjusts the transducer position which changes the displayed image, and back to sonographer looking at the monitor. Figure created using Inkscape version 0.92 (https://inkscape.org/).

Deep learning is a highly sophisticated pattern recognition methodology, well suited, in theory, to recognizing image appearance characteristics associated with diagnostic criteria. This is a key reason why deep learning is currently having an impact on radiology for automation of reading of computed tomography (CT), magnetic resonance imaging (MRI), X-ray and optical imaging[10–14]. Ultrasound imaging is unique in the level of high skill required to acquire *and* interpret imaging "on-the-fly". Obtaining a high-quality and informative diagnostic ultrasound image requires substantial expertise, a skill traditionally acquired over several years. In addition, in the case of obstetric ultrasound, image quality depends on maternal habitus, fetal movements and fetal position[7]. To improve sonographer accuracy, efficiency, and productivity, we need to further our understanding of the interaction between the sonographer, transducer and live image[15–18].

The PULSE (Perception Ultrasound by Learning Sonographer Experience) project is an interdisciplinary effort that aims to understand the entire obstetric imaging scanning process as a data science problem. For this purpose, we have developed a custom-built dedicated multi-modality ultrasound system that acquires the ultrasound scan and logs the sonographer actions by simultaneously recording full-length ultrasound video scans, eye-tracking data, transducer motion data, and audio of the sonographer speaking. Our system is deployed in a tertiary hospital clinic to capture data during routine obstetric ultrasound scanning for women attending the clinic for first, second and third trimester scans.

The aims of the paper were, first, to outline the PULSE methodology for the acquisition of data on sonographer perceptions and actions. To the best of our knowledge, the system, and dataset are unique. Second, we describe some of the deep learning image analysis models that have been built using data from the multiple perceptual cues. Third, we present lessons leant from the new field of ultrasound big data science. Fourth, we discuss some future technical research directions and possible new clinical applications of the emerging deep learning-based technology.

Methods

In this prospective data collection study, we record anonymized full-length obstetric ultrasound scan videos while tracking the actions of the sonographer. In the current paper we describe the data acquisition process of the PULSE study, explain how we interpret the multimodal data, and present preliminary clinically meaningful analyses.

Routine care settings. Recruitment began in May 2018. During the study period, as part of standard care, all women booked for maternity care in Oxfordshire, in the United Kingdom, are offered three routine ultrasound scans during pregnancy: in the first trimester at 11–13+6 weeks, which includes assessment for viability, gestational age assessment and an offer of aneuploidy screening by measurement of the nuchal translucency; a 20-week anomaly scan; and assessment of fetal growth at 36 weeks[19,20]. Ultrasound examinations are carried out by accredited sonographers, supervised trainees or fetal medicine doctors using standard ultrasound equipment. For quality assurance purposes, the stored images and the reliability of measurements are regularly assessed by a senior accredited sonographer using established quality criteria[21].

If an abnormality is suspected during a routine scan, the pregnant woman is referred for further evaluation at the Fetal Medicine Unit, Oxford University Hospitals National Health Services (NHS) Foundation Trust. Such scans are not included in the current study.

Pregnant women as participants. Pregnant women attending routine obstetric scans are invited to participate in the study by agreeing to have their ultrasound scan recorded by video. The inclusion criteria are age > 18 years old and the ability to provide verbal and written informed consent in English. Multiple gestations are not excluded. Following consent, each pregnant woman is given a unique study participant number to allow the anonymization of data.

The routine ultrasound scan is carried out without any modifications for the purpose of this study. Hence, ultrasound scans are performed according to national and unit guidelines. As for all women, scan results are given to the woman after the ultrasound examination and further management is performed according to national and local protocols by the appropriate healthcare professional. Women are not required to attend any follow-up or subsequent research scans. Women who consent to participate, but subsequently request to be withdrawn from the study, are excluded from the analysis.

Sonographers as participants. Participating sonographers are accredited sonographers, trainees supervised by accredited sonographers, or fetal medicine doctors. For the purpose of this study, we refer to all ultrasound operators, sonographers and fetal medicine doctors, as sonographers. The sonographer inclusion criteria are practicing obstetric ultrasound, and willing and able to give informed consent in English for participation in the study. Following written informed consent, each sonographer is assigned a unique study number. All participating sonographers are given an introduction about the purpose and aims of the study, followed by specific individual training as required.

Sonographers are requested to maintain the routine scanning procedures and are not required to alter their practice for the purpose of the study.

Data collection. *Ultrasound system.* All scans are performed using a commercial General Electric (GE) Healthcare Voluson E8 or E10 (Zipf, Austria) ultrasound machines equipped with standard curvilinear (C2-9-D, C1-6-D, C1-5-D) and 3D/4D transducers (RAB6-D, RC6M). The system is equipped with customized additions for recording scans, eye-tracking and transducer motion. Tracking of the vaginal transducer is not part of the study.

Video recording. The secondary video output from the ultrasound machine is connected to a computer equipped with a video grabbing card (DVI2PCIe, epiphany video, Palo Alto, California) and purpose-built software to ensure real-time anonymization of the video. Hence, the saved videos include no personal details. Full-length ultrasound scans are recorded using the ultrasound machine full high-definition (HD) resolution (1920 × 1080 pixels) at 30 frames per second. Video files are recorded using lossless compression.

Eye tracking. Eye tracking is achieved using a remote eye-tracker (Tobii Eyetracking Eye Tracker 4C, Danderyd, Sweden) mounted below the ultrasound machine display monitor. For each ultrasound frame, we record the exact point of sonographer gaze. The accuracy of the eye-tracking method in our setting was previously validated[22]. Sonographers do not have any visual or other signal to know that the eye-tracking device is functioning.

Transducer tracking. Transducer motion tracking is achieved using an Inertial Measurement Unit (IMU), a motion sensor, (NGIMU, X-IO Technologies, Bristol, UK) attached to the ultrasound transducer (Fig. 2). The three-axis accelerometer information is recorded at 100 Hz.

Audio recording. Sonographer voice recording is carried out using two microphones (PCC160, Crown HARMAN, Northridge, California). One microphone is located in proximity to the operator, next to the ultrasound machine display screen, to best capture the operator's voice. The second microphone is located away from the operator, next to the pregnant woman and any accompanying persons. This setup allows to isolate the sonographer's voice from that of others present in the scanning room. Differentiation of the operator's voice from that of others is carried out using the VoxSort Diarization software (Integrated Wave Technologies) and transcription is performed using the ELAN[23] annotation tool.

Sonographer interaction with ultrasound machine. Ultrasound machine keystrokes and cursor movement result in visual cues. Detection of the bidirectional sonographer-ultrasound machine communication is carried out with a custom-designed software. The software automatically processes the ultrasound video to deter-

Figure 2. Inertial Measurement Unit (IMU) device attached to a curvilinear ultrasound transducer, acting as the sonographer hand tracking device. Figure created using Microsoft Office version 365 (https://office.microsoft.com/).

mine sonographer manipulation of the ultrasound controls and machine-displayed values, allowing detection of events like "freeze", "save", or "clip save", and machine-displayed values such as thermal safety indices, or measurement values. The purpose-built software, which performs a frame-by-frame analysis, was implemented in Python (http://www.python.org, version 3.7.0) using OpenCV (http://www.opencv.org, version 3.4) and Tesseract (http://www.github.com/tesseract-ocr, version 3.05) libraries. Extracting the machine parameters displayed in the graphical user interface was achieved through pattern recognition and optical character recognition (OCR).

Sample size. Owing to the observational nature of the study, calculating a meaningful clinical sample size is challenging. For machine learning, it has been previously suggested that "the more the better"[24]. Based on logistical and clinical considerations, including an aim to achieve an adequate sample across different sonographers and across fetal gestational age, we prospectively chose to recruit a total of 3000 routine scans, with up to 20 sonographers performing these scans, to provide approximately 1000 scans for each gestational age group.

Data analysis. The data collected present holistic and very rich information about obstetric sonography. In order to analyze the data, the following strategies were used:

(1) Second-trimester workflow analysis using video understanding (Fig. 3). Since the second-trimester anomaly scan represents the most widely conducted obstetric ultrasound scan, we chose to analyze the workflow of these scans. *Partitioning*—initially, we automatically extracted 5-s clips from the videos that represent
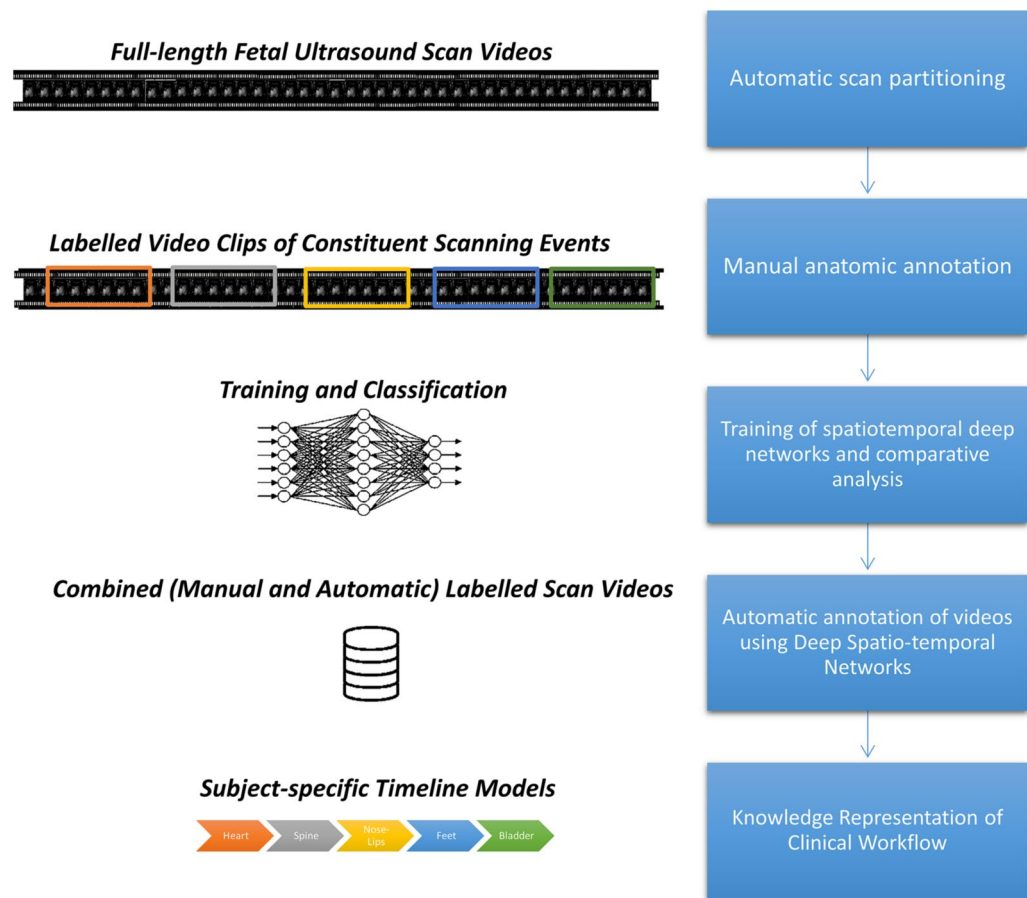
Figure 3. Outline of the clinical workflow analysis pipeline. Stepwise approach to second-trimester workflow analysis: (1) partitioning—automatically extraction of 5-s clips from the videos that represent important events; (2) manual labeling of the training dataset; (3) training a spatiotemporal deep network; (4) automatic labelling of the entire dataset; and (5) analysis of second-trimester scans as a sequence of organs being scanned. Figure created using Microsoft Office version 365 (https://office.microsoft.com/).

important events by automatic detection of video freeze, image save, or clip save. We defined these as important scan events because sonographers freeze the screen when concentrating on an acquisition and image/ clip save is carried out when the sonographer is satisfied with the acquisition displayed on the monitor.

Labeling—The very large number of 5-s clips makes it impractical to manually label all of them. Therefore, we developed a deep learning model to automate this process. We first carried out manual labeling of approximately 20% of the available video clips according to the anatomy/organ/standard plane: head–brain, face (axial, coronal, or sagittal), arms, hands, situs, thorax–heart, abdomen, umbilical cord abdominal insertion, genitalia, bladder, legs, femur, feet, spine, kidneys, full-body fetal sagittal view, placenta-amniotic fluid, maternal anatomy (like uterine artery), 3D/4D mode, mixed, and unidentified. Thereafter, we labeled approximately 300 s-trimester scans, using machine learning method[25,26]. In brief, we developed and trained deep spatiotemporal networks, after an exploratory analysis of several deep architectures such as 2D ConvNets, 3D ConvNets, Recurrent Neural Networks, model fusion and initialization techniques[25]. We comparatively evaluated the optimally-learned spatiotemporal deep neural network, used it to classify sequential events in unseen full-length ultrasound scan videos, and temporally regularized the predicted result.

At the final step, each second-trimester scan was represented as a continuous line of fetal standard plane (anatomy/organ) versus time. This chosen layout allows easy visualization and interpretation of the type, duration and order of the anatomy being evaluated.

(2) Motion: The inertial measurement unit (IMU), which is attached to the ultrasound transducer, was used to record the rotational rate and linear acceleration for each axis of the sensor. The rotational rate fused with additional sensory information (gravity, magnetic field) is fed into an attitude and heading reference system (AHRS) that computes the absolute orientation in 3D space. We analyzed the transducer movement to compute the three-dimensional orientation and linear acceleration of the transducer throughout full-length scans.

(3) Natural language of ultrasound: audio recordings were transcribed into text to build machine learning models for image captioning. These textual captions allowed studying the language sonographers use while
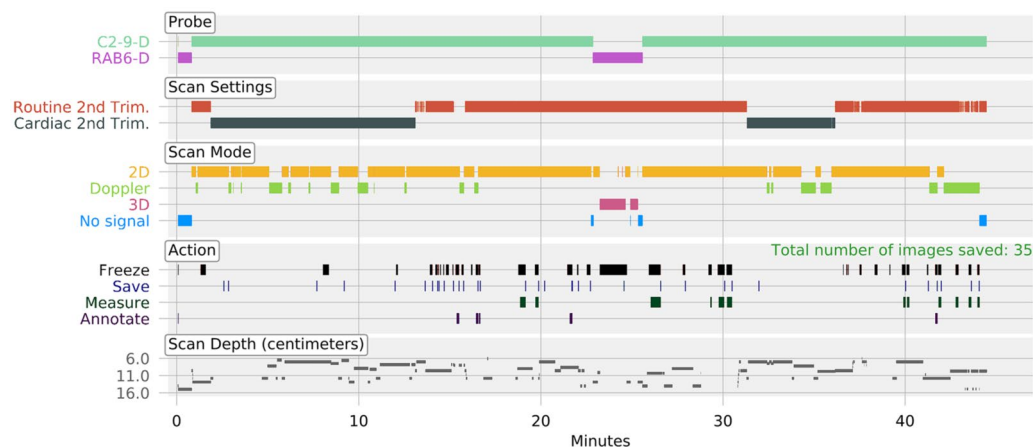
Figure 4. Anomaly scan machine parameters and sonographer actions according to time in one representative anomaly scan. A representative second-trimester anomaly scan. The overall duration of the scan was approximately 45 min, of which the majority of time was live scanning with the C2-9D probe. A large proportion of the scan was dedicated toward cardiac scanning and 35 episodes of image same were detected. Figure created using Python's Matplotlib library[53] and Microsoft Office version 365 (https://office.microsoft.com/).

performing a scan. Breakdown of the speech from a representative sample of scans determined that the distribution of adjectives, determiners, nouns, and verbs was 12.7%, 22.2%, 28.0%, and 16.0%, respectively[27]. The remaining 21.1% were prepositions, pronouns, adverbs, and other parts-of-speech.

(4) Eye-tracking: Using eye tracking we were able to determine the objects that a sonographer looks at. For instance, in the video clip 1 (Video S1), initially, it appears that the fetal profile is being examined; but the second part of the clip shows, by displaying the eye-tracking point on the video, that in fact the fetal brain is being studied (note that the green dot is added to the video to facilitate the presentation of our findings, and is not visible to a sonographer in the study).

Episodes of interest: For each scan video, the customized software detected episodes (periods of time) where a defined screen area was being looked at (such as the "measurement box" or the "bioeffect box"). The software detected uninterrupted fixations, defined as uninterrupted sonographer gaze toward the area of interest, lasting ≥100ms. If the fixation is interrupted, this was considered as a single episode of eye fixation if this interruption was 400 ms or less; or as a separate fixation if it is more than 400 ms[28,29]. We verified this choice of threshold by randomly looking at more than 50 detected fixations and making sure that the threshold resulted in no false positives.

Ethics approval. Ethics approval was granted by the West of Scotland Research Ethics Service, UK Research Ethics Committee (Reference 18/WS/0051). All methods were carried out in accordance with relevant guidelines and regulations. Written informed consent was obtained from all participants.

Results

Workflow analysis. A total of 341 full-length second trimester anomaly scans, performed by ten sonographers, with an average duration of 36.2 ± 11.6 min per scan, were available for analysis (65 K frames/anomaly scan), representing a total of approximately 205 h of video. Figure 4 shows the automatically extracted machine settings and sonographer actions in one typical full-length anomaly scan. In this representative anomaly scan, the overall duration was 44 min, of which 77% was in live-scanning, and a total of 35 images and clips were saved.

Out of the 341 scans, 62 anomaly scans were randomly selected and manually labeled (annotated) as the training dataset by a clinical obstetric ultrasound expert and seven image analysis scientists with knowledge of obstetric scanning. A high inter-annotator agreement (76.1%) was found between the annotators and a confusion matrix comparing the primary annotator and other annotators is displayed in Fig. 5A. We note that for a few labels, representing structures often seen together, the confusion was high—for example hands and face—as the fetal hands are often in proximity to the face, making both visible in the same image (Fig. 5A). To train a deep neural network, the 12 most commonly used labels were selected, representing 88% of the 5-s clips (Table 1). A supervised deep learning network architecture for automatic temporal semantic labelling performed the labelling of the 279 remaining un-labeled full-length anomaly scans. A sample of 28 scans (10%) from the automatically labelled dataset was randomly selected and manually labeled. Figure 5B presents a confusion matrix comparing the agreement between manual and automatic labeling which overall was found to be 76.4%. Statistical analysis of the manual and automatically labelled scans showed a high Pearson's correlation $\rho = 0.98$ (p < 0.0001). Also, there was relatively high disagreement between labels that often display more than one structure at the same time: placenta and maternal structures as well as kidneys and abdomen (Fig. 5B).

With all scans labelled, we ended up with 1,158,782 labelled video frames. This allowed the anomaly scans to be visualized by organ as a function of time (25 representative scans are shown in Fig. 6). From the time-based
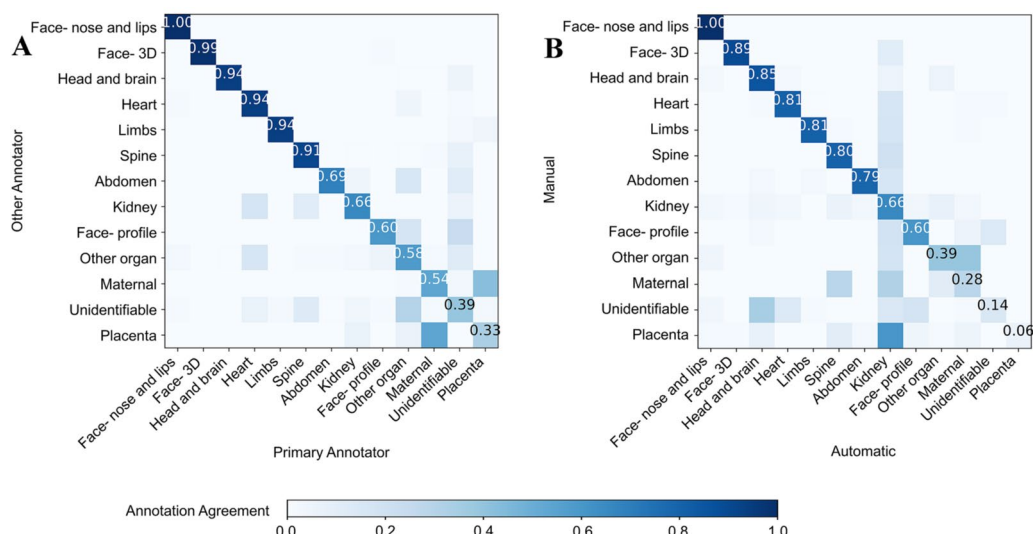
Figure 5. Confusion matrices for the comparison of scans labeling. (**A**) Scans labeled by different human annotators and (**B**) Scans manually labeled vs. those automatically labeled by a deep neural network. Figure created using the Python programming language (Python Software Foundation, https://www.python.org/).

Anatomical label	Prevalence (%)
Thorax–heart	19.6
Head–brain	11.5
3D/4D mode	11.2
Unidentified	11.0
Spine	6.1
Abdomen	5.5
Maternal anatomy	5.5
Coronal face	4.4
Placenta-amniotic fluid	4.2
Sagittal face (profile)	3.3
Kidneys	2.9
Femur	2.7

Table 1. Prevalence of the 12 most common anatomical labels in 62 full-length second-trimester anomaly scans.
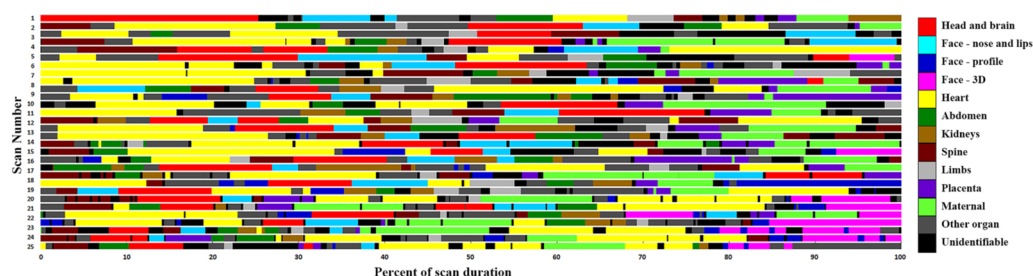

Figure 6. Anatomic plane being evaluated according to time (normalized to percent) in 25 representative anomaly scans. Figure created using MATLAB version 9.6 (https://www.mathworks.com/).

visualizations, on average, 21.5% (20.4–22.6%) of the scan length was dedicated to cardiac imaging, and 11.4% (10.8–12.0%) to the fetal head and brain (Table 2). It is also evident that on several scans, such as scan #25, the sonographer dedicated multiple episodes to scan the fetal heart.

Anatomical label	Average (95% CI) duration dedicated to anatomy
Thorax–heart	21.5% (20.4–22.6%)
Head–brain	11.4% (10.8–12.0%)
3D/4D mode	3.1% (2.6–3.6%)
Unidentified	17.7% (16.7–18.7%)
Spine	8.8% (8.0–9.5%)
Abdomen	5.9% (5.4–6.4%)
Maternal anatomy	10.5% (9.8–11.1%)
Coronal face	4.1% (3.7–4.4%)
Placenta-amniotic fluid	4.3% (3.9–4.8%)
Sagittal face (profile)	5.2% (4.7–5.8%)
Kidneys	1.6% (1.3–1.8%)
Femur	3.0% (2.7–3.3%)
Other organ	3.1% (2.4–3.8%)

Table 2. Proportion of scan duration dedicated toward each of the 12 most common anatomical labels in 341 full-length second-trimester anomaly scans. Data is percent (95% confidence interval).
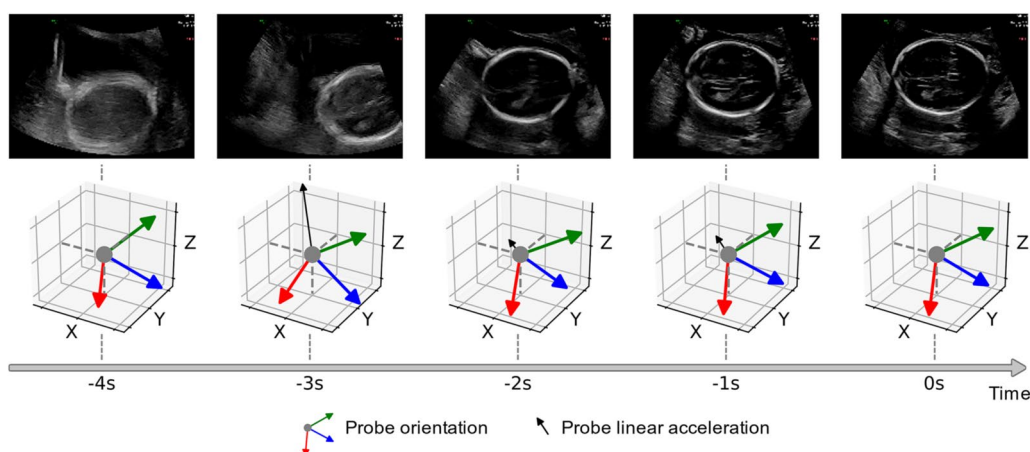

Figure 7. Example of transducer motion data. The figure displays the transducer motion data during the acquisition of the trans-ventricular (TV) standard biometry head and brain plane. The first row shows the ultrasound machine displayed frames (in 1-s intervals) before freezing (time = 0 s). The second row shows the corresponding transducer orientation and acceleration vector for each frame. Figure created using Python's Matplotlib library[53] and Inkscape version 0.92 (https://inkscape.org/).

Transducer motion. Figure 7 shows frames from a representative 5-s acquisition of the head trans-ventricular standard biometry plane and corresponding ultrasound transducer motion information in 1-s intervals. It is evident that the sonographer continuously manipulated the transducer position in order to acquire the desired imaging plane. The example shows the link between transducer orientation and linear acceleration and the ultrasound image. This motion information revealed how the sonographer arrived at the desired view, which is not implicitly seen from the video data by itself.

Image captioning. We empirically observe that operators use sentences that are up to 83 words long when describing ultrasound video with a vocabulary consisting of approximately 344 unique words. The small number of unique words in this sonographer vocabulary demonstrates that sonographers tend to speak in a very particular way to describe the content of ultrasound scans. Using this real-world sonographer vocabularies, we were able to build image captioning models to generate relevant captions for anatomical structures of interest that appear in routine fetal ultrasound scans[27]. Figure 8 compares a caption spoken by a sonographer with one generated by our automatic natural language processing (NLP) based image captioning method that models the vocabulary commonly used by operators. It is notable that the generated caption is similar to the operator spoken word in terms of sonographer linguistic style.

Safety of ultrasound. The Thermal Index is an indicator of the risk of tissue heating, displayed by standard ultrasound machines. To ensure safety, ultrasound sonographers are required to adhere to recommended ultra-
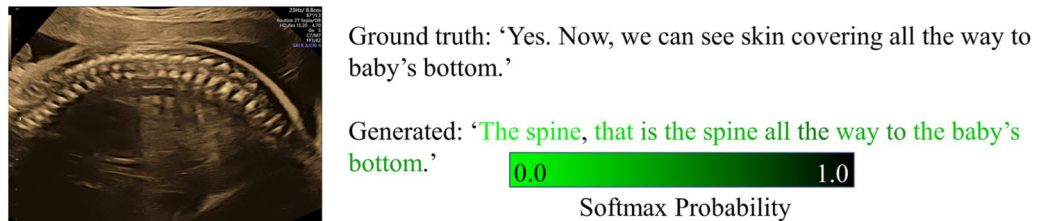
Figure 8. Image captioning: the ground truth and a generated caption are shown for the fetal spine. The darker the green color, the higher probability of the associated with a generated word (Softmax probability). Figure created using Python's Matplotlib library[53] and Microsoft Office version 365 (https://office.microsoft.com/).

sound thermal safety indices exposure times[30]. Using PULSE video and eye tracking data we can study how well sonographers adhere to the recommendations in practice. Specifically, software was implemented to automatically monitor the values displayed and gaze toward the "bioeffect box" displayed by the ultrasound machine. 17 sonographers performed 178, 216, and 243 ultrasound scans at the first, second and third trimesters, respectively. Gaze tracking showed that the displayed "bioeffect box" was looked at in only 27 of the 637 routine scans (4.2%). Despite the infrequent assessment of the thermal safety indices, the machine presets allowed for the recommended thermal safety indices to be kept during all routine scans and using the different ultrasound modes: B-mode, Color/Power Doppler, and Pulsed Wave Doppler[31].

Bias in biometric measurements. Fetal biometry is an essential part of the majority of pregnancy scans. It is the focus of the examination in the third trimester, and variability in fetal measurements is also higher at this gestation. Therefore, we analyzed a total of 272 third-trimester scans performed by 16 sonographers with associated gaze data that were available for analysis of the standard biometric measurements. During these scans, we automatically identified 1409 biometric measurements: 354 head circumference, 703 abdominal circumference, and 352 femur length. During biometry, ultrasound machines display the "measurement box" that displays the measurement result. Eye-tracking demonstrated that sonographers look at the ultrasound machine displayed values in > 90% of standard fetal biometric measurements. This bias causes the sonographers to "correct" the measurement by adjusting the caliper placement toward the expected measurement for the actual gestational age[32].

Discussion

In this paper, we demonstrate that ultrasound data science can offer fresh new clinical insights to understand routine real-world obstetric ultrasound screening imaging. We explain how, using multi-modal scanning data, we can build deep learning models of clinical sonography and image analysis. To be able to describe how expert sonographers perform a diagnostic study of a subject, we present a novel multi-cue data capture system that records the scan concurrently with tracking the actions of the sonographer, including interaction with the machine, sonographer eye and transducer movements. By measuring different sonographer actions or tasks, we further aim to understand better several topics such as how to effectively and efficiently initially train sonographers, monitor learning progress, and enhance scanning workflow. Our data can also help design deep learning applications for assistive automation technology as described elsewhere[33–36].

Understanding sonography in quantitative detail allows us to show clinically meaningful examples such as workflow analysis, image captioning, bias identification and thermal safety indices monitoring. The workflow of obstetric ultrasound is a serial task of non-ordered standard planes acquisition and interpretation. For each standard plane, the process starts with the detection of the main anatomical region, fine-tuning to the target structure followed by an optimal acquisition in the correct plane, verifying normal appearance, measuring as required and ultimately storing. We have presented a novel system that records an entire ultrasound scan and partitions it according to the standard plane/organ being temporally studied, which results in a quantitative visualization of plane/organ vs time. This allows us to study ultrasound workflow as information science. For example, we observe that sonographers spend a large proportion of the scan undertaking fetal cardiac assessment, yet there is a large variation between sonographers. This could potentially mean that a specific sonographer may be struggling with this part of the scan and might, therefore, benefit from further training. Thus, analysis and quantification of ultrasound scans can allow monitoring of learning, identification of weaknesses, auditing, and verifying scan completeness. Ultimately, this should improve the cost-effectiveness of scans. Previous work in ultrasound learning and workflow in non-obstetric ultrasound fields has resulted in the creation of automated algorithms to aid transducer guidance, automatic scanning, and image analysis[37,38]. In obstetrics, real-time automatic identification of standard planes may aid diagnosis and quality assurance[33,36,39–46].

Eye-tracking has revealed important sonographer behaviors, including the tendency to bias fetal biometry measurements. Such operator bias is important for research and clinical settings and require further prospective assessment to understand its source and impact. Another important lesson learned from eye-tracking is that sonographers do not adequately monitor ultrasound thermal safety indices. This means that algorithms should be specifically designed to adhere to safety recommendations and not just "learn" how to scan from sonographers. Additionally, automated assistance to sonographers in monitoring of thermal safety indices should be

implemented; ultrasound machines should not just report the safety parameters, but also be designed to actively avoid exceeding exposure times or alert the sonographer if the safety parameters are breeched.

Planned analyses. In our study, in addition to ultrasound video, we store real-time information of transducer motion and eye-tracking. This additional sensory data is particularly interesting as it has the potential to improve machine learning algorithms for obstetric ultrasound related tasks using expert knowledge that is not available from ultrasound video alone.

Ultrasound transducer guidance is already available commercially for adult echocardiography[47], and image-based guidance solutions have been proposed for instance for ultrasound-guided regional anesthesia[37]. The transducer tracking data captured in this study will lead to an understanding of how to best maneuver a transducer to acquire the appropriate planes, and in turn, allow the development of "guidance" of the transducer to the correct scan plane. Due to the highly variable anatomy that is dependent on fetal position, acquiring these skills usually requires years of training. An algorithm that guides the transducer to the correct position in real-time potentially has an enormous advantage over the current one-on-one teaching that includes a lot of human trial-and-error. An early report on our work on this topic is described in[48].

The tasks captured by sonographers during a routine fetal anomaly ultrasound scan are characterized by properties such as their order and their duration, which are not strictly constrained in the scanning protocol. Using deep learning on ultrasound video, we comprehensively analyze and quantify sonographer clinical workflow in terms of the type, duration and sequence of the tasks constituting a full-length scan. Next, we plan to attempt improving the accuracy of the sonographer and automatically generated image description. Thereafter, we aim to evaluate the utility of the sonographer tracking in improving the workflow assessment model.

Research and clinical potential. In this study, we explore how the latest ideas from deep learning can be combined with clinical sonography big data to develop deep models and understanding to inform the development of the next generation of ultrasound imaging capabilities and make medical ultrasound more accessible to non-expert clinical professionals.

Despite half of a century of ultrasound usage, most advances have been in sensors and electronic systems, leaving the sonographer behavior field largely unexplored. We attempt to bridge the gap between an ultrasound device and the user by employing machine learning solutions that embed clinical expert knowledge to add interpretation power. The hope is that taking this approach may provide a major step towards making ultrasound a more accessible technology to less-expert users across the world. It is widely known that a shortage of sonographers and radiologists are a major bottleneck to ultrasound access worldwide[49,50] and that ultrasound sonographers are at risk for burn-out, and musculoskeletal injuries that potentially limit their professional work. Sonographers who have been clinically active for decades probably scan in a manner that poses minimal risk for injury. By capturing the actions of such experts, we may determine important characteristics of scanning that may prevent sonographer injury.

Strengths. The novel design that allows us to understand human perceptual information by recording the sonographer eye and transducer movements during real-life scans, while recording the resulting ultrasound video, are major strengths of the study. We have already shown robust constant findings from capture and analysis of multiple scans and sonographer. This includes verification of the ability to detect important scan events, to extract data from the ultrasound image, and to accurately identify the sonographer gaze. Also, in light of emerging adoption of artificial intelligence in radiology[51,52], the ability to implement related methods into obstetric ultrasound is likely to progress the field of clinical obstetric sonography.

Limitations. Our study is not without limitations. Mainly, we present the study design and preliminary findings that require further in-the-field clinical validation. Also, before such an algorithm could be implemented in clinical practice, we aim to improve the inter-annotator and manual to automatic agreement accuracy. Currently, the relatively high agreement allows interpretation of data, but an algorithm to fit clinical settings, the agreement should be improved. Additionally, the limited number of sonographers in our ultrasound unit make it difficult to convincingly assess the differences between novice and expert sonographers. However, the need for thorough clinical validation is expected with any novel idea and we were able to show the preliminary findings are easily interpretable.

Conclusion

In conclusion, we have applied the latest ideas from deep learning to a vast amount of real-world clinical data from multiple perceptual cues, to describe and understand how sonographers scan and ultrasound images are acquired. We are currently using the derived deep learning models to design new assistive tools for ultrasound acquisition and image interpretation. These tools aim to maximize the diagnostic capabilities of ultrasound as well as improve the accuracy and reproducibility of sonography to contribute to advancing its use in the developed and developing world.

Data availability

The datasets analysed during the current study are not publicly available due to patient data governance policy. Analyses performed during the current study are available from the corresponding author on reasonable request.

References

1. United Kingdom National Screening Committee, P. H. E. Population screening programmes https://www.gov.uk/topic/population-screening-programmes (2020).
2. Abramowicz, J. S. Benefits and risks of ultrasound in pregnancy. *Semin. Perinatol.* **37**, 295–300. https://doi.org/10.1053/j.semperi.2013.06.004 (2013).
3. Pooh, R. K. & Kurjak, A. Novel application of three-dimensional HDlive imaging in prenatal diagnosis from the first trimester. *J. Perinat. Med.* **43**, 147–158. https://doi.org/10.1515/jpm-2014-0157 (2015).
4. Powers, J. & Kremkau, F. Medical ultrasound systems. *Interface Focus* **1**, 477–489. https://doi.org/10.1098/rsfs.2011.0027 (2011).
5. Abu-Rustum, R. S. & Abuhamad, A. Z. Fetal imaging: Past, present, and future. A journey of marvel. *BJOG* **125**, 1568. https://doi.org/10.1111/1471-0528.15343 (2018).
6. van Velzen, C. L. *et al.* Prenatal detection of congenital heart disease—Results of a national screening programme. *BJOG* **123**, 400–407. https://doi.org/10.1111/1471-0528.13274 (2016).
7. Benacerraf, B. R. *et al.* Proceedings: Beyond Ultrasound First Forum on improving the quality of ultrasound imaging in obstetrics and gynecology. *Am. J. Obstet. Gynecol.* **218**, 19–28. https://doi.org/10.1016/j.ajog.2017.06.033 (2018).
8. Qiu, X. *et al.* Prenatal diagnosis and pregnancy outcomes of 1492 fetuses with congenital heart disease: Role of multidisciplinary-joint consultation in prenatal diagnosis. *Sci. Rep.* **10**, 7564. https://doi.org/10.1038/s41598-020-64591-3 (2020).
9. Engelbrechtsen, L. *et al.* Birth weight variants are associated with variable fetal intrauterine growth from 20 weeks of gestation. *Sci. Rep.* **8**, 8376. https://doi.org/10.1038/s41598-018-26752-3 (2018).
10. Rajpurkar, P. *et al.* Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. (2017).
11. Choi, K. J. *et al.* Development and validation of a deep learning system for staging liver fibrosis by using contrast agent–enhanced CT images in the liver. *Radiology* **289**, 688–697 (2018).
12. McKinney, S. M. *et al.* International evaluation of an AI system for breast cancer screening. *Nature* **577**, 89–94. https://doi.org/10.1038/s41586-019-1799-6 (2020).
13. Ardila, D. *et al.* End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat. Med.* **25**, 954–961. https://doi.org/10.1038/s41591-019-0447-x (2019).
14. De Fauw, J. *et al.* Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat. Med.* **24**, 1342–1350. https://doi.org/10.1038/s41591-018-0107-6 (2018).
15. Coiera, E. The fate of medicine in the time of AI. *Lancet* **392**, 2331–2332. https://doi.org/10.1016/S0140-6736(18)31925-1 (2018).
16. Kalayeh, M. M., Marin, T. & Brankov, J. G. Generalization evaluation of machine learning numerical observers for image quality assessment. *IEEE Trans. Nucl. Sci.* **60**, 1609–1618. https://doi.org/10.1109/TNS.2013.2257183 (2013).
17. Choy, G. *et al.* Current applications and future impact of machine learning in radiology. *Radiology* **288**, 318–328. https://doi.org/10.1148/radiol.2018171820 (2018).
18. Noble, J. A., Navab, N. & Becher, H. Ultrasonic image analysis and image-guided interventions. *Interface Focus* **1**, 673–685. https://doi.org/10.1098/rsfs.2011.0025 (2011).
19. Drukker, L.. *et al.* How often do we incidentally find a fetal abnormality at the routine third-trimester growth scan? A population-based study *Am. J. Obstet. Gynecol.* **223**, 919.e1-919.e13. https://doi.org/10.1016/j.ajog.2020.05.052 (2020).
20. Kullinger, M., Granfors, M., Kieler, H. & Skalkidou, A. Discrepancy between pregnancy dating methods affects obstetric and neonatal outcomes: A population-based register cohort study. *Sci. Rep.* **8**, 6936. https://doi.org/10.1038/s41598-018-24894-y (2018).
21. Sarris, I. *et al.* Standardisation and quality control of ultrasound measurements taken in the INTERGROWTH-21st Project. *BJOG* **120**(Suppl 2), 33–37. https://doi.org/10.1111/1471-0528.12315 (2013).
22. Chatelain, P., Sharma, H., Drukker, L., Papageorghiou, A. T. & Noble, J. A. Evaluation of gaze tracking calibration for longitudinal biomedical imaging studies. *IEEE Trans. Cybern.* **50**, 153-163 https://doi.org/10.1109/TCYB.2018.2866274 (2020).
23. Sloetjes, H. & Wittenburg, P. In *6th international Conference on Language Resources and Evaluation (LREC 2008).*
24. Figueroa, R. L., Zeng-Treitler, Q., Kandula, S. & Ngo, L. H. Predicting sample size required for classification performance. *BMC Med. Inform. Decis. Mak.* **12**, 8. https://doi.org/10.1186/1472-6947-12-8 (2012).
25. Sharma, H. *et al.* Spatio-temporal partitioning and description of full-length routine fetal anomaly ultrasound scans. *Proc IEEE Int. Symp. Biomed. Imaging* **16**, 987–990. https://doi.org/10.1109/ISBI.2019.8759149 (2019).
26. Sharma, H. *et al.* Knowledge representation and learning of operator clinical workflow from full-length routine fetal ultrasound scan videos. *Med. Image Anal.* **69**, 101973. https://doi.org/10.1016/j.media.2021.101973 (2021).
27. Alsharid, M. *et al.* Captioning ultrasound images automatically. *Med. Image Comput. Comput. Assist. Interv.* **22**, 338–346. https://doi.org/10.1007/978-3-030-32251-9_37 (2019).
28. Brysbaert, M. Arabic number reading: On the nature of the numerical scale and the origin of phonological recoding. *J. Exp. Psychol. Gen.* **124**, 434–452. https://doi.org/10.1037/0096-3445.124.4.434 (1995).
29. Salvucci, D. D. & Goldberg, J. H. In *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications* 71–78 (Association for Computing Machinery, 2000).
30. Safety Group of the British Medical Ultrasound Society. Guidelines for the safe use of diagnostic ultrasound equipment. https://www.bmus.org/static/uploads/resources/BMUS-Safety-Guidelines-2009-revision-FINAL-Nov-2009.pdf. (2009).
31. Drukker, L., Droste, R., Chatelain, P., Noble, J. A. & Papageorghiou, A. T. Safety indices of ultrasound: Adherence to recommendations and awareness during routine obstetric ultrasound scanning. *Ultraschall. Med.* **41**, 138–145. https://doi.org/10.1055/a-1074-0722 (2020).
32. Drukker, L., Droste, R., Chatelain, P., Noble, J. A. & Papageorghiou, A. T. Expected-value bias in routine third-trimester growth scans. *Ultrasound Obstet. Gynecol.* **55**, 375-382 https://doi.org/10.1002/uog.21929 (2019).
33. Cai, Y., Sharma, H., Chatelain, P. & Noble, J. A. Multi-task SonoEyeNet: Detection of fetal standardized planes assisted by generated sonographer attention maps. *Med. Image Comput. Comput. Assist. Interv.* **11070**, 871–879. https://doi.org/10.1007/978-3-030-00928-1_98 (2018).
34. Droste, R. *et al.* Ultrasound image representation learning by modeling sonographer visual attention. *Inf. Process Med. Imaging* **26**, 592–604. https://doi.org/10.1007/978-3-030-20351-1_46 (2019).
35. Jiao, J., Droste, R., Drukker, L., Papageorghiou, A. T. & Noble, J. A. Self-supervised representation learning for ultrasound video. *Proc. IEEE Int. Symp. Biomed. Imaging* **1847–1850**, 2020. https://doi.org/10.1109/ISBI45749.2020.9098666 (2020).
36. Cai, Y. *et al.* Spatio-temporal visual attention modelling of standard biometry plane-finding navigation. *Med. Image Anal.* **65**, 101762. https://doi.org/10.1016/j.media.2020.101762 (2020).
37. Smistad, E. *et al.* Automatic segmentation and probe guidance for real-time assistance of ultrasound-guided femoral nerve blocks. *Ultrasound Med. Biol.* **43**, 218–226. https://doi.org/10.1016/j.ultrasmedbio.2016.08.036 (2017).
38. Shin, H. J., Kim, H. H. & Cha, J. H. Current status of automated breast ultrasonography. *Ultrasonography* **34**, 165–172. https://doi.org/10.14366/usg.15002 (2015).
39. Chen, H. *et al.* Standard plane localization in fetal ultrasound via domain transferred deep neural networks. *IEEE J. Biomed. Health Inform.* **19**, 1627–1636. https://doi.org/10.1109/JBHI.2015.2425041 (2015).

40. Baumgartner, C. F. *et al.* SonoNet: Real-time detection and localisation of fetal standard scan planes in freehand ultrasound. *IEEE Trans. Med. Imaging* **36**, 2204–2215. https://doi.org/10.1109/TMI.2017.2712367 (2017).
41. SonoScape. S-Fetus. http://www.sonoscape.com/html/2018/exceed_0921/86.html. (2020).
42. Xie, H. *et al.* Using deep learning algorithms to classify fetal brain ultrasound images as normal or abnormal. *Ultrasound Obstet. Gynecol.* **56**, 579-587 https://doi.org/10.1002/uog.21967 (2020).
43. Yaqub, M. *et al.* OP01.10: Auditing the quality of ultrasound images using an AI solution: ScanNav˚ for fetal second trimester ultrasound scans. *Ultrasound Obstet. Gynecol.* **54**, 87–87. https://doi.org/10.1002/uog.20656 (2019).
44. Yaqub, M., Kelly, B., Papageorghiou, A. T. & Noble, J. A. Guided random forests for identification of key fetal anatomy and image categorization in ultrasound scans. in: Navab N. Hornegger J. Wells W. Frangi A. Medical image computing and computer-assisted intervention—MICCAI 2015. 687–694 (Springer International Publishing) (2015).
45. Yaqub, M. *et al.* Plane localization in 3-D fetal neurosonography for longitudinal analysis of the developing brain. *IEEE J. Biomed. Health Inform.* **20**, 1120–1128. https://doi.org/10.1109/JBHI.2015.2435651 (2016).
46. Burgos-Artizzu, X. P., Perez-Moreno, A., Coronado-Gutierrez, D., Gratacos, E. & Palacio, M. Evaluation of an improved tool for non-invasive prediction of neonatal respiratory morbidity based on fully automated fetal lung ultrasound analysis. *Sci. Rep.* **9**, 1950. https://doi.org/10.1038/s41598-019-38576-w (2019).
47. FDA Authorizes Marketing of First Cardiac Ultrasound Software That Uses Artificial Intelligence to Guide User. https://www.fda.gov/news-events/press-announcements/fda-authorizes-marketing-first-cardiac-ultrasound-software-uses-artificial-intelligence-guide-user. (2020).
48. Droste, R., Drukker, L., Papageoghiou, A. & Noble, J. Automatic probe movement guidance for freehand obstetric ultrasound. *Med Image Comput Comput Assist Interv.* **12263**, 583-592. https://doi.org/10.1007/978-3-030-59716-0_56 (2020).
49. Shah, S. *et al.* Perceived barriers in the use of ultrasound in developing countries. *Crit. Ultrasound J.* **7**, 28. https://doi.org/10.1186/s13089-015-0028-2 (2015).
50. Waring, L., Miller, P. K., Sloane, C. & Bolton, G. Charting the practical dimensions of understaffing from a managerial perspective: The everyday shape of the UK's sonographer shortage. *Ultrasound* **26**, 206–213. https://doi.org/10.1177/1742271X18772606 (2018).
51. Drukker, L., Noble, J. A. & Papageorghiou, A. T. Introduction to artificial intelligence in ultrasound imaging in obstetrics and gynecology. *Ultrasound Obstet. Gynecol.* **56**, 498-505 https://doi.org/10.1002/uog.22122 (2020).
52. Soffer, S. *et al.* Convolutional neural networks for radiologic images: A radiologist's guide. *Radiology* **290**, 590–606. https://doi.org/10.1148/radiol.2018180547 (2019).
53. Hunter, J. D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **9**, 90–95. https://doi.org/10.1109/MCSE.2007.55 (2007).

Acknowledgements

Author contributions

L.D.: planning, carrying out, analysing and writing. H.S.: planning, carrying out, analysing and writing. R.D.: planning, carrying out, analysing and writing. M.A.: planning, carrying out, analysing and writing. P.C.: planning, carrying out, analysing and writing. J.A.N.: planning, carrying out, analysing and writing. A.T.P.: planning, carrying out, analysing and writing.

Funding

Competing interests

JAN and ATP are Senior Scientific Advisors of Intelligent Ultrasound Ltd. LD, HS, RD, MA, PC declare that they have no competing interests as defined by Nature Research, or other interests that might be perceived to influence the results and/or discussion reported in this paper.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-92829-1.

Correspondence and requests for materials should be addressed to J.A.N.

Reprints and permissions information is available at www.nature.com/reprints.

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