



# City Research Online

## City St George's, University of London

**Citation:** Matthew, Jacqueline, Skelton, Emily, Day, Thomas G, Zimmer, Veronika A, Gomez, Alberto, Wheeler, Gavin, Toussaint, Nicolas, Liu, Tianrui, Budd, Samuel, Lloyd, Karen, et al (2022). Exploring a new paradigm for the fetal anomaly ultrasound scan: Artificial intelligence in real time. *Prenatal Diagnosis*, 42(1), pp. 49-59. doi: 10.1002/pd.6059

This is the published version of the paper.




This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

**Permanent repository link:** <https://openaccess.city.ac.uk/id/eprint/28933/>

**Link to published version:** <https://doi.org/10.1002/pd.6059>

**Copyright and Reuse:** Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to [City Research Online policy](#).

# Exploring a new paradigm for the fetal anomaly ultrasound scan: Artificial intelligence in real time

Jacqueline Matthew<sup>1,2</sup>  | Emily Skelton<sup>1,2,3</sup>  | Thomas G. Day<sup>1,2</sup>  |  
 Veronika A. Zimmer<sup>1</sup> | Alberto Gomez<sup>1</sup> | Gavin Wheeler<sup>1</sup> | Nicolas Toussaint<sup>1</sup> |  
 Tianrui Liu<sup>4</sup> | Samuel Budd<sup>4</sup> | Karen Lloyd<sup>1</sup> | Robert Wright<sup>1</sup> | Shujie Deng<sup>1</sup> |  
 Nooshin Ghavami<sup>1</sup> | Matthew Sinclair<sup>1</sup> | Qingjie Meng<sup>4</sup> | Bernhard Kainz<sup>4</sup> |  
 Julia A. Schnabel<sup>1</sup> | Daniel Rueckert<sup>4,5</sup> | Reza Razavi<sup>1,2</sup> | John Simpson<sup>1,2</sup> |  
 Jo Hajnal<sup>1</sup>

<sup>1</sup>School of Biomedical Engineering and Imaging Sciences, King's College London, London, UK

<sup>2</sup>Guy's and St Thomas' NHS Foundation Trust, London, UK

<sup>3</sup>School of Health Sciences, City University of London, London, UK

<sup>4</sup>Department of Computing, Imperial College London, London, UK

<sup>5</sup>School of Informatics and Medicine, Klinikum Rechts der Isar, Technical University of Munich, Munich, Germany

## Correspondence

Jacqueline Matthew, Perinatal Imaging and Health Department, King's College London, St Thomas Hospital, Westminster Bridge Rd, London SE1 7EH, UK.

Email: [Jacqueline.matthew@kcl.ac.uk](mailto:Jacqueline.matthew@kcl.ac.uk)

## Funding information

National Institute for Health Research; Wellcome Trust, Grant/Award Number: 102431; Wellcome/EPSRC Centre for Medical Engineering, Grant/Award Number: WT203148/Z/16/Z

[Correction added on 28 October 2021, after first online publication: The expansion for FASP is included and the figure 4 is updated. Throughout the article text, foetal/foetus is replaced with fetal/fetus.]

## Abstract

**Objective:** Advances in artificial intelligence (AI) have demonstrated potential to improve medical diagnosis. We piloted the end-to-end automation of the mid-trimester screening ultrasound scan using AI-enabled tools.

**Methods:** A prospective method comparison study was conducted. Participants had both standard and AI-assisted US scans performed. The AI tools automated image acquisition, biometric measurement, and report production. A feedback survey captured the sonographers' perceptions of scanning.

**Results:** Twenty-three subjects were studied. The average time saving per scan was 7.62 min (34.7%) with the AI-assisted method ( $p < 0.0001$ ). There was no difference in reporting time. There were no clinically significant differences in biometric measurements between the two methods. The AI tools saved a satisfactory view in 93% of the cases (four core views only), and 73% for the full 13 views, compared to 98% for both using the manual scan. Survey responses suggest that the AI tools helped sonographers to concentrate on image interpretation by removing disruptive tasks.

**Conclusion:** Separating freehand scanning from image capture and measurement resulted in a faster scan and altered workflow. Removing repetitive tasks may allow more attention to be directed identifying fetal malformation. Further work is required to improve the image plane detection algorithm for use in real time.

Jacqueline Matthew, Emily Skelton and Thomas G. Day are joint first authors.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. Prenatal Diagnosis published by John Wiley & Sons Ltd.

### Key points

#### What is already known about this topic?

- Artificial intelligence has shown great promise in medical diagnosis, including in antenatal settings
- Most published work has been based on retrospective data, with very little work exploring how AI might be used in real-life clinical practice

#### What does this study add?

- We have shown that real time use of AI in obstetric ultrasound scanning is feasible and can fundamentally disrupt how sonographers perform the scan
- AI-assisted scans were significantly faster than standard manual scans
- Automatically measured fetal biometry was highly accurate
- The performance of automatic standard plane acquisition needs to be improved before these tools can enter mainstream clinical use

## 1 | INTRODUCTION

The mid-trimester fetal anomaly ultrasound (US) scan is now in widespread use, but quality is not consistent. Despite a trend showing improvement, international antenatal screening detection rates remain variable. For example, the rate of antenatal diagnosis of severe congenital heart disease internationally ranges from 13%–87%, with wide variation within countries.<sup>1–3</sup> A recent study examining why fetal heart defects are missed during screening found that in the majority of such cases, either the correct sonographic plane was not correctly obtained, or the defect was clearly demonstrated on screen but not recognised by the operator.<sup>4</sup>

Artificial intelligence (AI) has been shown to achieve human-level performance in some medical imaging analysis tasks.<sup>5</sup> This raises the potential for automating aspects of the fetal US scan, including automated image identification in real time and automatically measured fetal biometry.<sup>6–13</sup> Although excellent model performance is often described using retrospective data *in silico*, relatively little work has been published on translating these findings to the more chaotic and unpredictable clinical world. This has created a gap between AI development in the literature and peer-reviewed clinical validation in real-world clinical settings.<sup>14</sup> Very few prospective clinical trials have compared AI to human performance in medical imaging, and this pilot study bridges that gap by embedding integrated AI tools into a live clinical scan.<sup>15</sup>

Although commercial US manufacturers have recently introduced new AI-driven obstetric ultrasound products to the market, these aim to complement and augment the sonographer's skills as 'assistive' optional compartmentalised tools and do not fundamentally alter how the scan is performed.<sup>16</sup> We propose a new paradigm, in which AI is used to free the sonographer from scan processes that force task switching. This would fundamentally alter the scanning workflow and may allow the human operator to focus more attentively on other diagnostic aspects of the examination during live scanning.

The aim of the present study was to pilot the end-to-end automation of multiple elements of the mid-trimester US screening scan

using AI-enabled tools and to assess the impact of these on the efficiency and quality of the fetal examination compared to a standard manual scan.

## 2 | METHODS

This study was performed as part of the intelligent Fetal Imaging and Diagnosis (iFIND) project, NRES number = 14/LO/1806 (ISRCTN = 16542843). All participants gave informed written consent. Scans were performed by two UK trained obstetric sonographers with 20 years combined experience (JM and ES) and equivalent professional qualifications. The trial design was a single centre prospective method comparison study. A sample size target of between 20 and 25 participants was pragmatically selected to assess the feasibility of end-to-end automation and to assess the initial impact on efficiency, measurement reliability and image quality before a larger trial.

### 2.1 | Participants

The inclusion criteria were: completed 18<sup>+0</sup>–20<sup>+6</sup> routine fetal anomaly ultrasound scan with no subsequent onward referral to specialist services, and gestational age between 20<sup>+0</sup> and 24<sup>+0</sup> weeks.

### 2.2 | Study protocol

Participants were required to attend a dedicated research clinic, where two fetal US scans were performed sequentially, using a Philips EpiQ US system with a C5-1 MHz curvilinear transducer (Philips Healthcare, Best, Netherlands). The manual (standard) fetal ultrasound scan was performed, and the AI-assisted scan (aided by automated scan plane detection, automated biometry and autoreport tools) was conducted as a comparator modality (see Figure 1 for the scan protocol flowchart). Both scans were performed in line with the UK's national FASP Fetal Anomaly Screening Programme guidelines

(considered the gold standard). Supporting Information S1 gives further details of the research US scanning methodology with example standard images. The 13 standard planes for which acquisition was attempted during each scan were: trans-ventricular brain; trans-cerebellar brain; abdominal circumference view; femur length view; facial profile; lips and nose; right and left outflow tract; four chamber view; three vessel tracheal view; kidneys; sagittal spine; coronal spine. The sonographers alternated between performing the manual scan and the AI-assisted scan for each participant to control for any difference in scanning ability or technique, blinded to each other's scan procedure and results. The time to perform each scan was recorded independently. The scan was considered completed when each view was either successfully acquired or had been attempted a maximum of three times (after the mother was repositioned, or drank some water to encourage fetal movement, as is common practice in a clinical setting).

The same display was used for both manual scanning and AI-assisted scanning. In manual mode only the information normally displayed by the EpiQ system was presented. In the AI-assisted mode, a 'traffic light' system (Figure 2) provided additional real-time information for each anatomical view about the automated image capture and storage (green: high confidence of detection; amber: moderate confidence; red: insufficient confidence). The only input data for the AI algorithms was from the full live stream 2D US data that was collected for the duration of the examination. During AI-assisted examinations, the sonographer dynamically scanned and observed the fetus until they were satisfied that a comprehensive visual assessment had been completed. This was in combination with confirmation that the required planes had been captured (i.e. green 'traffic lights'). During this scan, the automated tools capture the 13 standard views and the fetal biometry of the biparietal diameter (BPD), head circumference (HC), abdominal circumference (AC) and

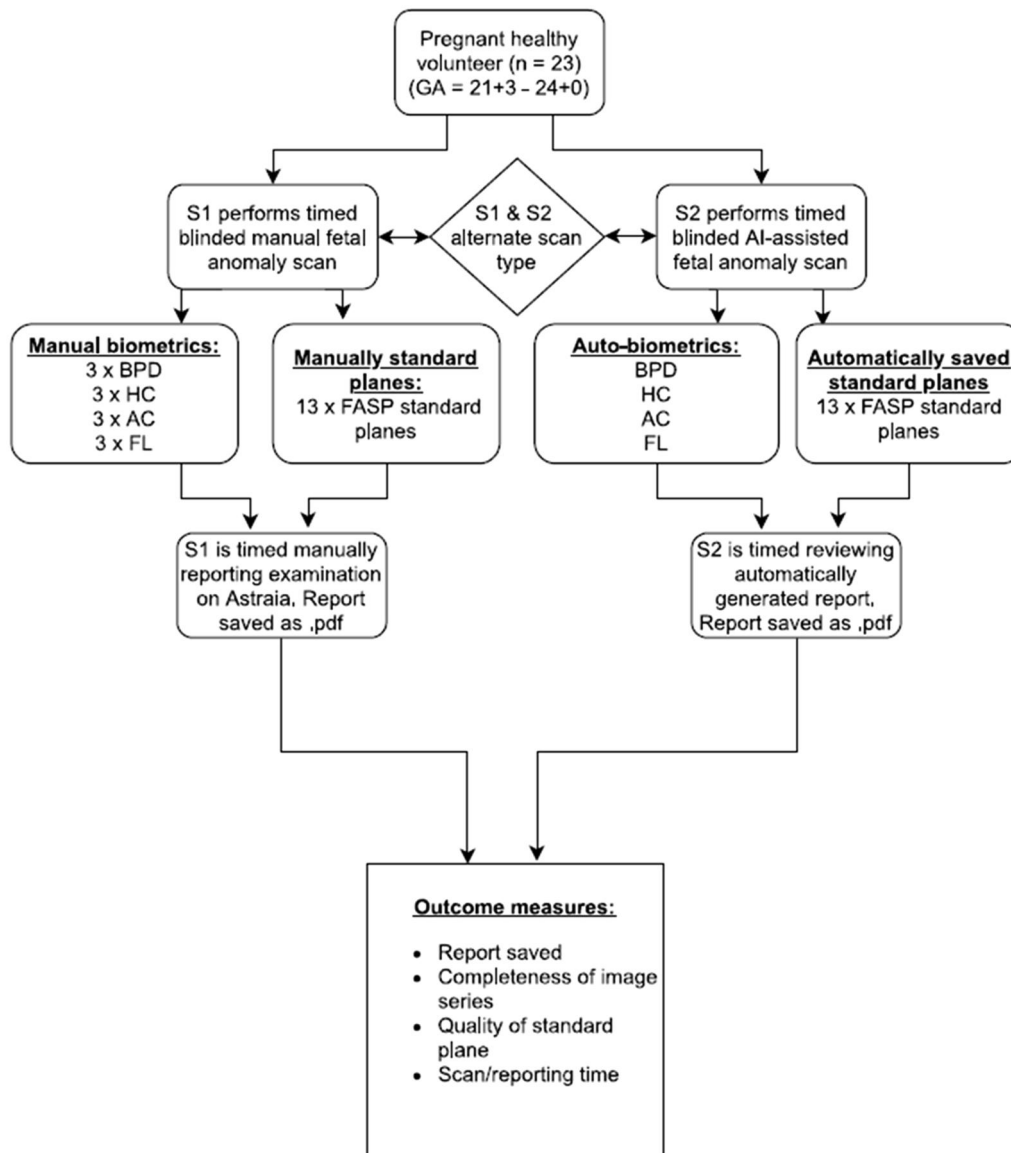
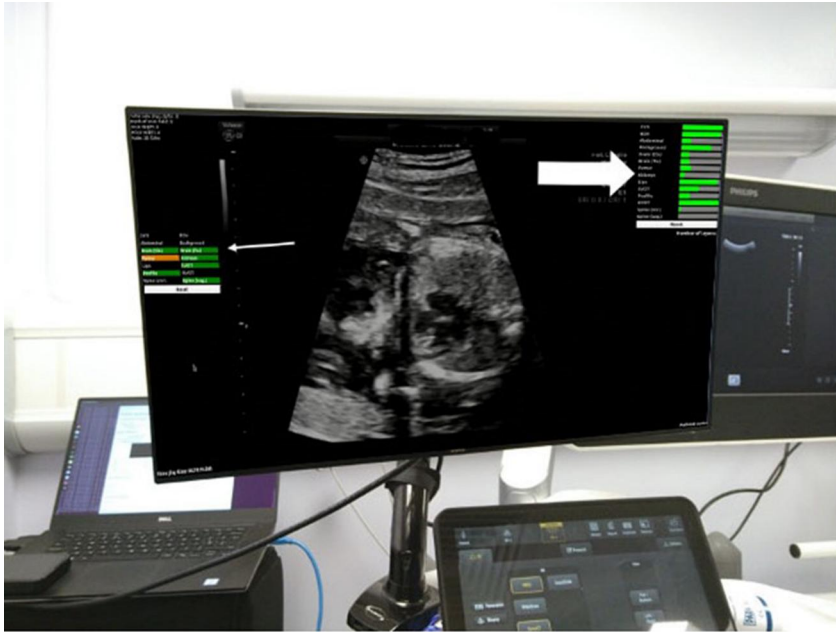


FIGURE 1 Research scan protocol flowchart



**FIGURE 2** Research clinic room set up and display monitors. Large white arrow: AI feedback overlay, displaying real-time detection confidence for each standard view. Small white arrow: 'Traffic light' system indicating the overall confidence of the completeness of the data capture for each standard view (high, moderate, low) [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

femur length (FL) via a standard laptop (DELL XPS15 Intel® Core™ i7-6700HQ CPU @ 2.60 GHz × 8, GeForce GTX 960M 2 GB, RAM 16 GB with Ubuntu Linux 18.04LTS). The laptop was connected to the digital video output of the US system via a frame grabber (Epiphan DVI2USB3.0).

At the end of an AI-assisted scan, a report (automatically populated with five candidate images for each required plane) was generated in .html format (Supporting Information S2). To complete the auto-report, the sonographer selected the best quality plane from each set of candidate views. This selection triggered the inclusion of the corresponding biometric measurements automatically extracted from the images, presented both as numerical values and on relevant growth curves. Other clinical data (e.g. gestational age and fetal position) were manually entered by the sonographer, and the report was saved as a .pdf file when complete.

During the manual scan, acquisition of each standard view was attempted, and biometric measurements were repeated, three times on different images (as is considered best practice in the FASP guidelines).<sup>17</sup> The 13 standard views (which include the four core biometric views) were saved in the standard way on the ultrasound machine's hard drive ready to be sent to a picture archiving and communication system (PACS) system after the examination. The written manual report was produced using Astraia, a standard obstetric reporting software package (Astraia software GmbH, Munich, Germany). This required physical input of the technically best biometric measurement selected by the sonographer.

### 2.3 | Image quality and biometric measurement validation

After the data collection period of the study was complete, a third independent sonographer (KL), blinded to the scan method,

performed a subjective assessment of image quality. A purpose-built interface that presented side-by-side images of the same nominal standard view from the manual and the AI-assisted scans unlabelled and in random order was used for this assessment. For each image pair, the sonographer was asked to select which was of a higher quality, or if they were of equal quality. To allow for inter-observer comparison of biometry, the same sonographer also repeated all biometric measurements offline using MITK workbench software (German Cancer Research Center Division of Medical Image Computing, Heidelberg, Germany).

### 2.4 | Analysis of constituent parts of the scan

To validate our findings from the blinded and paired comparison study between the manual and AI scanning methodologies, we used a bespoke software script to automatically analyse the frozen, that is 'non-scanning' time in the manual scans for each study participant. To further assess how scan time savings may translate into an uncontrolled clinical environment, 782 consecutive scan datasets from the previously described iFIND-1 cohort<sup>7</sup> which had been verified for completeness, were analysed in a similar fashion.

### 2.5 | Artificial intelligence algorithms

To ensure a seamless real-time end user experience during the AI-assisted scan, the following six elements were required to work simultaneously (see Table 1, and for full technical methods see Supporting Information S3):

- (1) Real-time image capturing through HDMI video output and frame grabber.

- (2) Automatic standard plane detection, image saving, and clustering<sup>7</sup> (SonoNet (EpiQ), ClusterSPD.v2).
- (3) Automatic biometric extraction<sup>6,18</sup> (BrainNet.v1, AbdoNet.v2, FemurNet.v1).
- (4) Automatic report generation (Autoreport.v2).
- (5) Real-time display of the original images and AI-derived information ('traffic light' system).
- (6) Dedicated computer running all the above simultaneously in real time.

The number of standard plane image frames collected per case by the SonoNET (EpiQ) algorithm was too large for a sonographer to review during clinical reporting. 'ClusterSPD' was developed to address this and was used to provide five candidate views from which the operator performed a final offline selection. This provided both an efficient and manageable process and afforded sonographers the chance for a second review of key clinical information.

## 2.6 | Data analysis

Means and standard deviations (SD) were calculated for the biometry, scanning and reporting times for each scan condition. Paired absolute and relative difference were compared between scan conditions across subjects and tested using a two-tailed paired *t*-test (where  $p < 0.05$  was considered significant). The growth in gestational days equivalent to size differences in different measurements were calculated using fetal growth reference charts.<sup>19-22</sup> All data were analysed using SPSS (version 24, SPSS Inc, Chicago, Ill, USA) and Excel (version 15.0, Microsoft Corp, Redmond, Washington, USA). At the end of the study, a short feedback survey was conducted with the scanning sonographers to gain some structured feedback on the use of the tools.

## 3 | RESULTS

Between May 2019 and March 2020, 23 pregnant women were recruited into this single centre study to undergo both AI-assisted and standard manual prospective mid-trimester fetal US scans as per the NHS FASP protocol (Figure 1).<sup>17</sup> In total, 299 manually acquired and 260 AI-assisted FASP standard image views were achieved. Table 2 outlines the baseline participant information, and Figure 3 demonstrates exclusions and losses from the final data used in analyses.

### 3.1 | Examination duration and processes

Scan times were 34.7% shorter in the AI-assisted scan compared to the manual scan (mean duration 14.32 vs. 21.93 min, mean time saving 7.62 min; Table 3). With the exception of two cases, AI-assisted scans were faster than the matched manual scan. When a

subgroup of manual scans was analysed, the mean duration for the screen to be frozen during the manual scans (corresponding to periods where the images were being visually assessed, measured, and saved) was 7.8 min, corresponding to 38.8% of the total scan duration (Table 4, where comparable data is also shown from a large clinical cohort database collected from a working antenatal clinic, further details in discussion).

Sonographers reported that the AI tools were easy to use, perceiving a change in their scanning approach during the AI-assisted scan compared to the manual scan. Both sonographers agreed that the AI tools made it easier to concentrate on image interpretation, with no adverse impact on patient interaction during the scan (Figure 4).

Despite a fundamental change in written report production (during which the AI-assisted process required the sonographers to undertake an additional review of the automatically obtained images for selection into the report: see Supporting Information S2), there was no significant difference in the time taken to complete the reports between the two methods (Table 3). When selecting a specific image for use, the auto-report automatically included any corresponding biometric measurements, displaying these in conventional tabular form and plotting them on population centile charts.

### 3.2 | Scan completeness, image quality and biometrics

In considering completeness of the reports for the core fetal views (trans-ventricular brain view (TV), trans-cerebellar brain view (CB), abdominal circumference view (AC) and femur length view (FL)), the AI-assisted report included 93% of the required views, compared to 98% in the manual report (Figure 5 and Table 5). When all 13 standard views were considered (i.e. core biometry views plus additional facial profile, lips and nose, right and left outflow tract, four chamber view of the heart, three vessel tracheal view, kidneys, sagittal and coronal spine: see Supporting Information S1), the AI-assisted report successfully saved 73% of the required images compared to 98% in the manual report.

Independent, blinded assessment of image quality demonstrated that the automatically extracted image was of a superior quality than the paired manual equivalent image in 33%, 41%, 44%, and 19% of the TV, CB, AC, and FL view respectively, and the manual view superior in 33%, 29%, 38%, and 69%. The views were felt to be of equivalent quality in 33%, 29%, 19%, and 13% (Figure 5).

There were no statistically significant differences between the manual and AI-assisted measurements for any of the biometrics other than for the head circumference (Table 6). Although for this measurement there was a statistically significant difference between the manual and automated measurements, the human inter-observer difference was greater than the manual versus AI-assisted difference (−5.37 mm, SD 3.36 vs. −2.44 mm, SD 3.65). Additionally, neither of these differences were considered clinically significant (i.e. they

TABLE 1 Algorithm data and clinical outcome measures

	Algorithm name and version	Input data type	Output data selection	Poor/missing data handling	Output data
Four fetal measurements and intermodality variation	BrainNet, AbdoNet, FemurNet	Complete ultrasound real-time video feed	ClusterSPD/human interaction	All poor or missing data noted in autoreport	Autoreport: biometry values and growth charts
Thirteen standard views completeness and quality	SonoNet (EpiQ)	Complete ultrasound real-time video feed	ClusterSPD/human interaction	All poor or missing data noted in autoreport	Autoreport: Standard views saved

TABLE 2 Baseline participant information

Category	Sub-category	Result
Number of participants, <i>n</i>		23
GA median (range), weeks and days		23 <sup>+2</sup> (21 <sup>+4</sup> -24 <sup>+0</sup> )
BMI median (range), kg/m <sup>3</sup>		23.4 (19.5-30.7)
Postnatal outcome, <i>n</i>	Livebirth	19
	Perinatal death	0
	Missing data (delivered elsewhere)	4
Fetal position, <i>n</i>	Cephalic	13
	Breech	7
	Transverse	3
Placental position, <i>n</i>	Posterior	14
	Anterior	6
	Fundal or lateral	3
Sonographer reported quality, <i>n</i>	Good	15 (3 scans restricted by fetal position)
	Satisfactory	8 (2 scans restricted by fetal mobility, 6 scans restricted by fetal position)
	Poor	0

Abbreviations: BMI, body mass index; GA, gestational age.

would not change routine clinical management) based on the associated equivalent change in gestational age (in this case, between 1 and 3 days).

## 4 | DISCUSSION

Standard manual obstetric US examinations feature an interrupted mode of operation with repeated pauses to save images and make measurements. A reporting process then follows, which is primarily a mechanical data entry task. In contrast, the AI-assisted approach that we tested fundamentally changed the way in which the scan was performed and reported. Many tasks were completely automated, enabling the sonographer to focus on optimal scanning and offering parent-centred care. The reporting process was transformed into a secondary image review task: as the images for review had been automatically acquired during live scanning, the operator could critically appraise image quality and evaluate likelihood of anomalies.

### 4.1 | Algorithm performance and clinic workflow

The AI-assisted scans were significantly faster (saving on average 34.7%) than manual scans. When we analysed the constituent parts of a subgroup of manual scans, we found that the screen was frozen for 39% of the total scan time (i.e. a similar duration to the AI time saving). Selecting the ideal image plane and placing callipers for biometric measures are not required in the AI-assisted scans, therefore it is likely that this is where the time savings are gained. To investigate how this may translate into a true clinical environment, we analysed the constituent parts of a retrospective sample of 782 full video recordings of anomaly scans from a single centre (Table 4). During these the proportion of the scan time where the screen was frozen was approximately 19.5%. This represents a clinically useful real-world time saving should our results be replicated outside of the controlled study environment.

Cognitive load theory concerns the overloading of working memory by extraneous factors that do not contribute to task completion.<sup>23</sup> Previous work demonstrates that AI assistance can

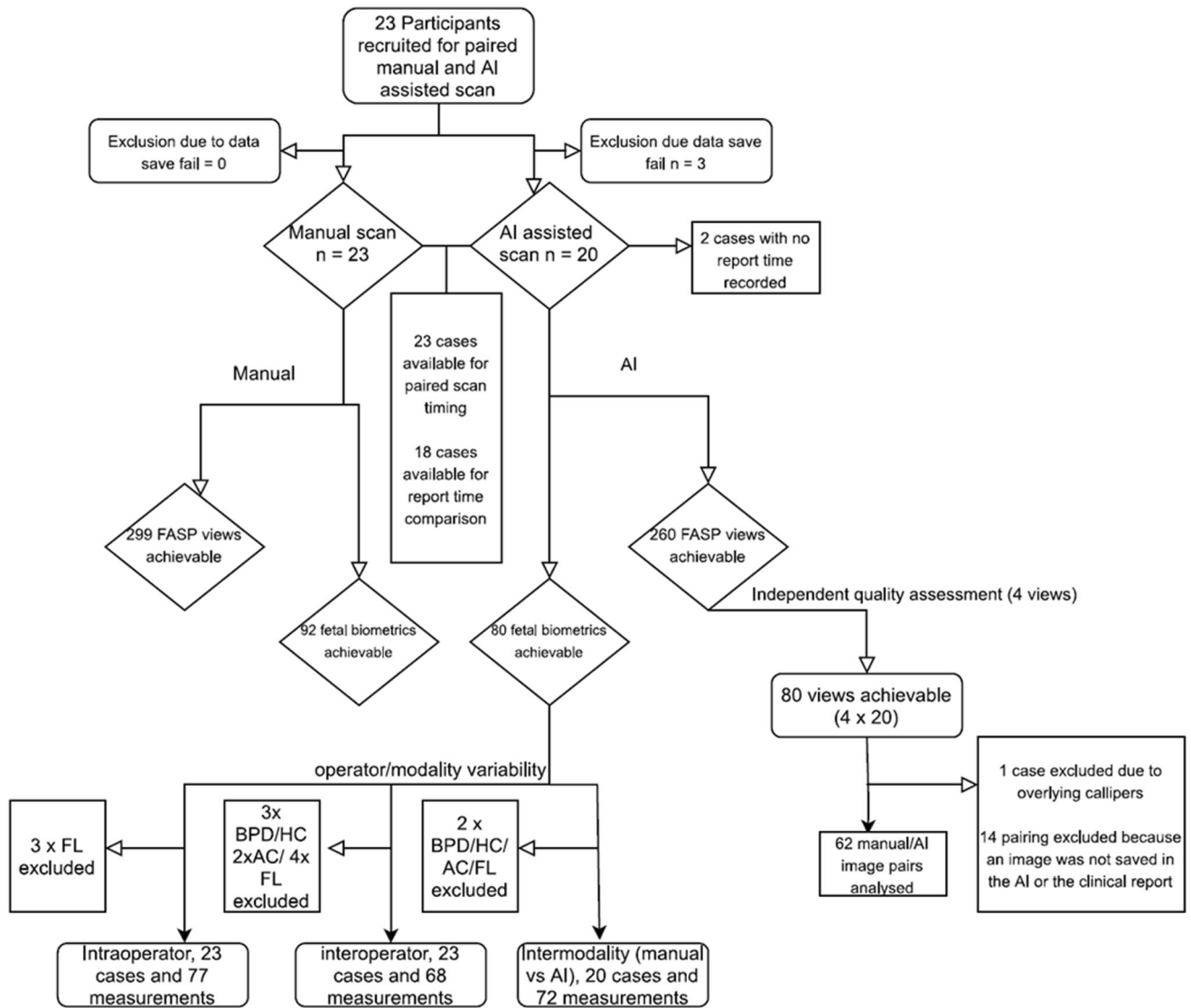


FIGURE 3 Participant and image data flowchart

TABLE 3 Scanning and reporting times for both manual and AI-assisted scans with mean paired difference

Timing type (n)	Method	Mean time (range), min	SD, min	Paired mean difference, min	p value
Scan time (20)	Manual	21.93 (17.33–29.28)	3.24	7.62	<0.0001
	AI-assisted	14.32 (10.93–22.53)	2.85		
Reporting time (18)	Manual	2.39 (1.63–4.67)	0.87	–0.09	Not sig.
	AI-assisted	2.49 (1.22–4.6)	0.76		

Abbreviations: min, time in minutes; SD, standard deviation.

TABLE 4 Proportion of frozen scan time from a ‘manual’ scan from pilot data and a clinical dataset

	Number of scans analysed	Mean total scan time (SD), min	Mean total frozen time (SD), min	Mean % of scan frozen (SD)	Mean number of freeze episodes (SD)
Manual scan from this study	14	20.1 (3.9)	7.8 (3.7)	38.8 (16.9)	34 (8)
Scans from previous large clinical cohort	782	19.6 (6.6)	3.8 (1.7)	19.4 (5.1)	51.7 (19)

Abbreviations: %, percentage; SD, standard deviation.



FIGURE 4 Sonographer feedback survey of using AI-tools [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

reduce cognitive load and improve performance in US-based tasks.<sup>24,25</sup> We did not directly assess cognitive load in this study, however sonographer feedback suggests that the less disrupted approach of the AI-assisted scan seemed more efficient and made it easier to focus on the interpretation (rather than acquisition) of relevant images.

We utilised a modified SonoNet algorithm<sup>7</sup> to automatically acquire standard planes (SonoNET EpiQ). This algorithm had excellent performance when previously tested with retrospectively acquired data, but unsurprisingly had poorer performance in our 'real-world' prospective setting. Considering the possible complete 13-image

dataset, a suitable image was saved for 73%. This is improved to 93% when we include only the four core views, but this is still inferior to the standard manual scan (98% for both). This discrepancy highlights an area for necessary improvements before this technology enters mainstream clinical use.

Automatically acquired fetal measurements were found to be highly reliable, and comparable with manually obtained measurements. No clinically significant differences were found between the two methods, and the discrepancy between automatic and manual measurements was less than the difference between two independent human observers. This demonstrates that as well as being faster to

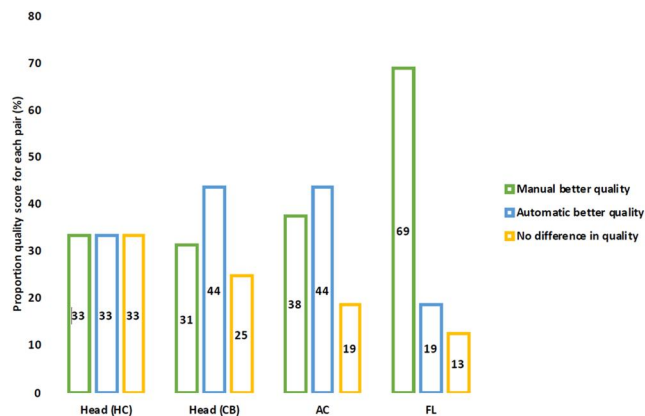


FIGURE 5 Proportional paired quality score by independent review [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

TABLE 5 Completeness of manual and AI-assisted scan views

	Report type	Number of cases, n	Total views possible, n	Successfully saved images, n (%)
Core views (TV, CB, AC, FL)	Manual	23	92	90 (98)
	AI-assisted	20	80	74 (93)
All 13 FASP views	Manual	23	299	294 (98)
	AI-assisted	20	260	191 (73)

Abbreviations: %, percentage; AC, abdominal circumference view; CB, trans-cerebellar brain view; FASP, fetal anomaly screening programme; FL, femur length view; TV, trans-ventricular brain view.

TABLE 6 Comparison of intra-observer, inter-observer, and inter-modality mean paired differences, with gestational age equivalent

	Mean (SD) paired difference, mm	GA equivalent of mean paired difference, days	p value
<b>Biparietal diameter</b>			
Intra-observer	-0.12 (1.07)	<1 day	0.59
Inter-observer	-1.26 (1.28)	2–3 days	<0.001
Manual versus AI-assisted	0.22 (2.44)	<1 day	0.704
<b>Head circumference</b>			
Intra-observer	1.47 (1.93)	<1 day	<0.001
Inter-observer	-5.37 (3.36)	3 days	<0.001
Manual versus AI-assisted	-2.44 (3.65)	1–2 days	0.011
<b>Abdominal circumference</b>			
Intra-observer	-1.09 (4.63)	<1 day	0.27
Inter-observer	-0.37 (4.3)	<1 day	0.7
Manual versus AI-assisted	2.81 (8.98)	1–2 days	0.2
<b>Femur length</b>			
Intra-observer	0.1 (0.85)	<1 day	0.607
Inter-observer	0.41 (0.95)	1–2 days	0.075
Manual versus AI-assisted	0.24 (1.79)	<1 day	0.575

Note: Significant results highlighted in bold.

Abbreviations: GA, gestational age; mm, millimeters; SD, Standard deviation.

acquire, automatic biometry measurement has human-level accuracy and repeatability consistent with technical validation experiments.<sup>6</sup>

## 4.2 | Challenges and limitations

Despite the recent volume of research on the use of AI in medical imaging, the majority of studies are based on the testing of trained algorithms using retrospectively acquired data.<sup>5,15</sup> Compared to other modalities, US has little work published on the use of AI.<sup>16,26</sup> This is likely due to additional challenges posed by ultrasound, particularly the high degree of operator-dependence and susceptibility to artefact. The way in which US is performed is also fundamentally different to other modalities, with images being acquired

and interpreted simultaneously, meaning that any useful AI system must work in real time. Fetal US is further challenged by other factors, including fetal movement, difficult acoustic windows, and very small physical sizes of anatomical structures. However, US imaging is accessible, relatively inexpensive, and portable. For these reasons, clinically useful AI-enabled US procedures have an unparalleled impact potential, both for the democratisation of technology for less expert clinical staff, and for imaging-based diagnosis and interventions in pregnancy.

Limitations of this study include a relatively small sample size, although this is to our knowledge the largest study of the prospective use of fully integrated AI in a complete fetal screening examination. Also, the fact that sonographers were unblinded to the knowledge that the subjects had previously had a routine anomaly scan without onward referral meant that they were aware of a low risk for significant clinical findings. The dedicated research setting included additional observers in the scan room (a timekeeper and engineers on standby for technical support), and the scan environment was artificial by clinical standards with a stricter protocol, possibly resulting in higher protocol adherence compared to a typical US clinic with multiple operators.

### 4.3 | Future research

The two areas requiring improvement before our package of AI tools could be implemented in a real-world clinical environment are: (1) the completeness of the AI-assisted scans for an extended set of standard views, and (2) quality of the acquired images. The quality of the acquired images compared between the AI-assisted and the manual scans (as assessed by a blinded independent observer) was broadly similar for the core views of the head and abdomen but were poorer in the views of the femur. These areas of potential improvement highlight the previously discussed challenges in translating AI into an actual clinical environment.<sup>14</sup>

This study was designed to pilot the introduction of AI into the UK fetal anomaly screening programme (FASP). As such, only views which are specified in the FASP protocol were used. There are additional views such as views of the fetal bladder or umbilical cord that are often used by fetal medicine specialists to diagnose important fetal disease. Also, more advanced ultrasound methods such as colour Doppler are routinely used by fetal medicine and cardiac specialists. Although not currently part of population-level screening in the UK, the application of AI to aid the interpretation of these additional views and other imaging techniques may be a novel way of improving the performance of the screening programme in the future.

## 5 | CONCLUSIONS

We have demonstrated the feasibility and utility of integrated AI-tools in the performance of obstetric US scans. These tools have the potential to transform the current practice ultrasound scanning methodology by altering workflow. By removing tasks that interrupt live

scanning, the ultrasound operator achieves a faster scan and with similar image quality, although further work is required to improve the performance of the image plane detection algorithms for some standard views when used in real time. The subjective sonographic experience was also changed, most likely a result of the reduction in task switching and interruption of observational scanning. Large scale multicentre trials motivated by smaller evaluation studies (such as this one) will be needed to determine if AI assistance results in increased antenatal fetal anomaly detection rates. We hypothesise that in the future, AI will release human operators from repetitive tasks, allowing them to focus on other aspects of the scan which could enhance scanning quality and antenatal care for expectant families.

### ACKNOWLEDGEMENTS

Firstly, we would like to thank the parent volunteers for taking part in this study. We would also like to thank the wider iFIND team and specifically Dr Silvia Giampieri, Petra Priebojova and Stefanie Chan for their administrative assistance and support in the conduct of this work. This work was supported by the Wellcome Trust (IEH Award, 102431), by core funding from the Wellcome/EPSCRC Centre for Medical Engineering (WT203148/Z/16/Z) and by the National Institute for Health Research (NIHR) Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust and King's College London and the NIHR Clinical Research Facility. Jacqueline Matthew and Thomas G. Day are supported by NIHR Doctoral Fellowships. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care. The funding bodies had no influence in the study design, data collection, analysis, interpretation, preparation of the manuscript, or decision to submit or publication.

### CONFLICT OF INTERESTS

AG reports consultancy fees from Ultromics Ltd. Bernhard Kainz reports consultancy fees from ThinkSono Ltd, Ultromics Ltd. and Cydar Medical Ltd. Daniel Rueckert reports consultancy fees from Heartflow, and IXICO, and in addition Daniel Rueckert has a patent US20200027237A1 pending. All other authors report no conflict of interests.

### ETHICS STATEMENT

The study has been granted NHS R&D and ethics approval, NRES ref no = 14/LO/1086.

### DATA AVAILABILITY STATEMENT

The collected and analysed datasets during the current study are available from the corresponding author upon request. The source code for the plug-in based real-time ultrasound software (PRETUS) is available at [github.com/gomezalberto/pretus](https://github.com/gomezalberto/pretus).

### ORCID

Jacqueline Matthew  <https://orcid.org/0000-0003-4754-0322>

Emily Skelton  <https://orcid.org/0000-0003-0132-7948>

Thomas G. Day  <https://orcid.org/0000-0001-8391-7903>

## REFERENCES

1. Quartermain MD, Pasquali SK, Hill KD, et al. Variation in prenatal diagnosis of congenital heart disease in infants. *Pediatrics*. 2015;136:e378-e385.
2. Bakker MK, Bergman JEH, Krikov S, et al. Prenatal diagnosis and prevalence of critical congenital heart defects: an international retrospective cohort study. *BMJ Open*. 2019;9:1-12.
3. National Institute for Cardiovascular Outcomes Research. *National Congenital Heart Disease Audit, 2019 Summary Report*. National Institute for Cardiovascular Outcomes Research, 2019.
4. Nisselrooij AEL, Teunissen AKK, Clur SA, et al. Why are congenital heart defects being missed? *Ultrasound Obstet Gynecol*. 2019;55(6):747-757.
5. Liu X, Faes L, Kale AU, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *Lancet Digit Health*. 2019;1:e271-e297.
6. Sinclair M, Baumgartner CF, Matthew J, et al. Human-level performance on automatic head biometrics in fetal ultrasound using fully convolutional neural networks. In: *Proceedings of Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS 2018*; 2018:714-717.
7. Baumgartner CF, Kamnitsas K, Matthew J, et al. SonoNet: real-time detection and localisation of fetal standard scan planes in freehand ultrasound. *IEEE Trans Med Imaging*. 2017;36:2204-2215.
8. Salim I, Cavallaro A, Ciofolo-Veit C, et al. Evaluation of automated tool for two-dimensional fetal biometry. *Ultrasound Obstet Gynecol*. 2019;54:650-654.
9. Sobhaninia Z, Rafei S, Emami A, et al. Fetal ultrasound image segmentation for measuring biometric parameters using multi-task deep learning. In: *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*. Institute of Electrical and Electronics Engineers Inc., 2019:6545-6548.
10. Cai Y, Sharma H, Chatelain P, Noble JA. Multi-task SonoEyeNet: detection of fetal standardized planes assisted by generated sonographer attention maps. In: *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Springer, 2018:871-879.
11. van den Heuvel TLA, de Bruijn D, de Korte CL, Ginneken B van. Automated measurement of fetal head circumference using 2D ultrasound images. *PLoS ONE*. 2018;13:e0200412.
12. Espinoza J, Good S, Russell E, Lee W. Does the use of automated fetal biometry improve clinical work flow efficiency? *J Ultrasound Med*. 2013;32:847-850.
13. Burgos-Artizzu XP, Coronado-Gutiérrez D, Valenzuela-Alcaraz B, et al. Evaluation of deep convolutional neural networks for automatic classification of common maternal fetal ultrasound planes. *Sci Rep*. 2020;10:1-12.
14. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med*. 2019;17:1-9.
15. Nagendran M, Chen Y, Lovejoy CA, et al. Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies in medical imaging. *BMJ*. 2020;368:1-12.
16. Drukker L, Noble JA, Papageorgiou AT. Introduction to artificial intelligence in ultrasound imaging in obstetrics and gynecology. *Ultrasound Obstet Gynecol*. 2020;56:498-505.
17. NHS Screening Programmes. *Fetal Anomaly Screening: Programme Handbook*. NHS Screening Programmes, 2015.
18. Sinclair M, Martinez JC, Skelton E, et al. Cascaded transforming multi-task networks for abdominal biometric estimation from ultrasound. In: *1st Conference on Medical Imaging with Deep Learning*. Amsterdam, 2018.
19. Chitty LS, Altman DG, Henderson A, Campbell S. Charts of fetal size: 2. Head measurements\*. *BJOG An Int J Obstet Gynaecol*. 1994;101:35-43.
20. Chitty LS, Altman DG, Henderson A, Campbell S. Charts of fetal size: 3. Abdominal measurements. *BJOG An Int J Obstet Gynaecol*. 1994;101:125-131.
21. Chitty LS, Altman DG, Henderson A, Campbell S. Charts of fetal size: 4. Femur length. *Br J Obstet Gynaecol*. 1994;101:132-135.
22. Loughna P, Chitty L, Evans T, Chudleigh T. Fetal size and dating: charts recommended for clinical obstetric practice. *Ultrasound*. 2009;17:161-167. doi:10.1179/174313409X448543
23. Paas F, van Gog T, Sweller J. Cognitive load theory: new conceptualizations, specifications, and integrated research perspectives. *Educ Psychol Rev*. 2010;22:115-121.
24. Aldekhyl S, Cavalcanti RB, Naismith LM. Cognitive load predicts point-of-care ultrasound simulator performance. *Perspect Med Educ*. 2018;7:23-32.
25. Buettner R. *Cognitive workload of humans using artificial intelligence systems: towards objective measurement applying eye-tracking technology*. 2013:37-48.
26. Rezazade Mehrizi MH, van Ooijen P, Homan M. Applications of artificial intelligence (AI) in diagnostic radiology: a technography study. *Eur Radiol*. 2021;31:1805-1811.

## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

**How to cite this article:** Matthew J, Skelton E, Day TG, et al. Exploring a new paradigm for the fetal anomaly ultrasound scan: Artificial intelligence in real time. *Prenat Diagn*. 2022;42(1):49-59. <https://doi.org/10.1002/pd.6059>