

Artificial Intelligence in Obstetrics

*Who's been making the decisions, what are they,
and where are they taking us?*

Trish Chudleigh

retired Research Sonographer

University of Cambridge & The Rosie Hospital, Cambridge

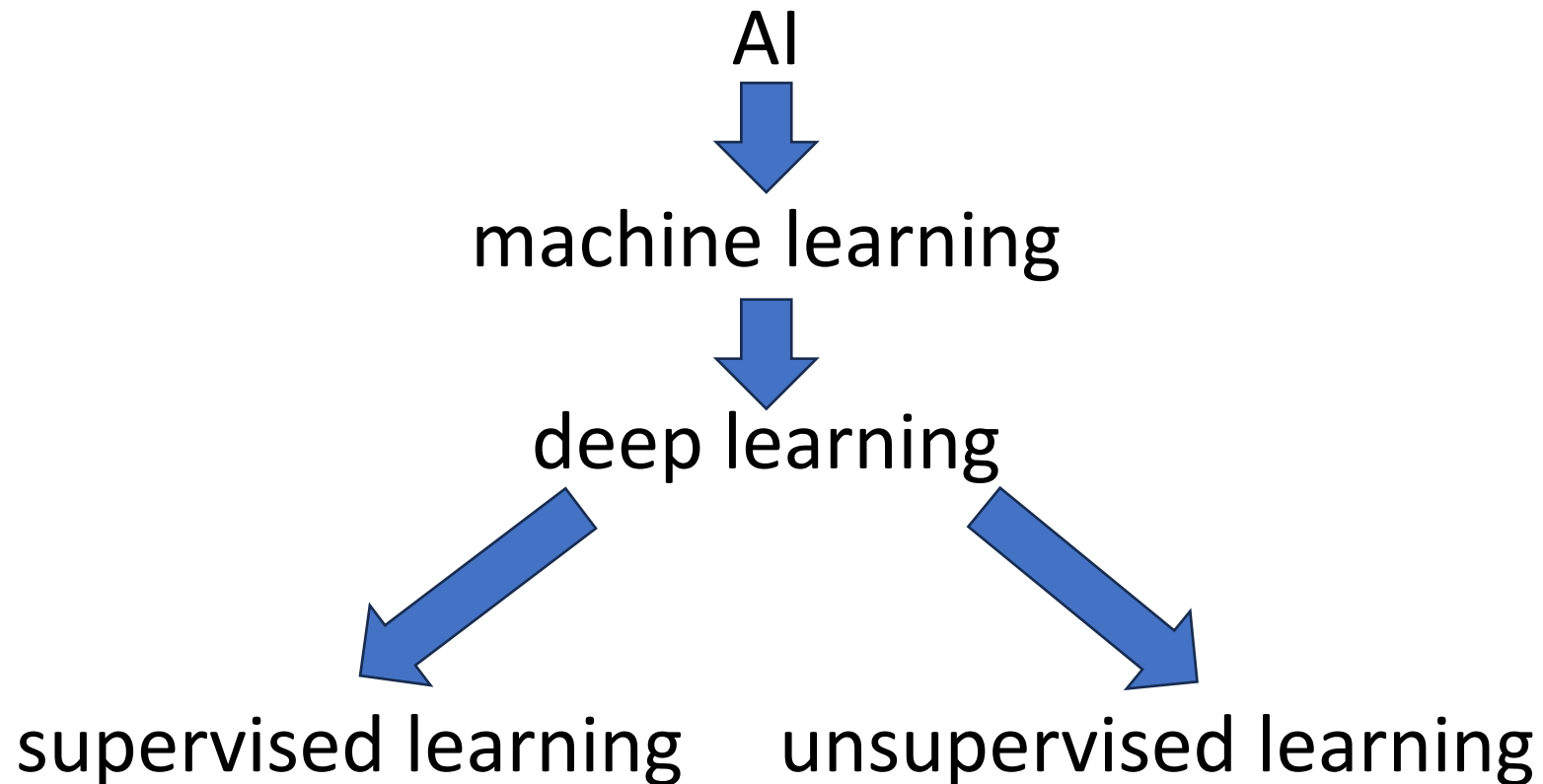
BMUS Study Day, 07 September 2024

What is Artificial Intelligence (AI)?

- 1950 – Alan Turing introduced the ‘Turing test’ – machine has passed the test if its evaluator cannot distinguish whether the intelligent behaviour has been demonstrated by a machine or a human
- 1955 – John McCarthy proposed term ‘artificial intelligence’
- Computers (computational algorithms), e.g. the calculator, follow sequence of rules & perform same function every time – ‘if this is the input, then that is the output’
- AI algorithm learns rules from training data presented to it
 - *1977 - Deep Blue outmatched Gary Kasparov world chess champion*
 - *2016 – AlphaGo defeated 9-dan Lee Sedol in ancient Chinese game of Go*

What is AI?

Describes tasks performed by machines or software that would normally require human brain power to accomplish



Machine Learning

- Ability to learn from data without being explicitly programmed to do so
- Statistical method that gradually improves as exposed to more data by extracting patterns from that data
e.g. Amazon shopping
Google searching

Deep Learning

Type of machine learning in which input & output are connected by hidden connections – artificial or convolutional neural networks (CNNs)

- CNNs learn from analysing very large datasets
 - *2 major classes – supervised & unsupervised*
 - *processes which occur inside the layers of the CNN are hidden, creating a 'black box'*
- DL performs especially well in pattern recognition within data - so great potential for clinical imaging applications

Supervised Learning

- CNN trained using labelled dataset
 - *annotating (labelling) training datasets performed manually - time consuming & expensive*
- CNN labelled dataset then evaluated using test dataset which contains unlabelled data
- Resulting output (e.g. normal/abnormal; fetal head or placenta) will have a prediction accuracy of the question being asked which must then be validated

Unsupervised Learning

- Training process that requires no labelling
- Saves time consuming, labour intensive & expensive human image labelling
- CNN learns from clustering scans that look similar to one another (brain/brain) or different from each other (fetal head/placenta)

Where are we now with AI?



AI is the teammate that clinicians need to make informed decisions faster, and that administrators need to meet increased demand, and support ultrasound users' well-being.






Consider:

- Imaging
- Biometry
- Setting & agreeing the standards
- Workflow

AI in the department – quality across the service

A full suite of AI-powered solutions

Our wide array of AI tools can help with many standard tasks including (but not limited to):

Guidance Onscreen guidance provides prompts to guide probe movements helping new users capture diagnostic-quality cardiac images	Standardizing AI tools can standardize some portions of obstetrical exams, improving efficiency by 65% ^v	Measuring Select cardiovascular measurements are made quickly and with just one click thanks to AI ^{vi}	Populating Automatically populating TI-RADS [™] descriptors makes thyroid cancer risk-assessment more accurate leading to a 57% reduction in benign biopsies ^{vii}	Labeling AI-powered automatic labeling of the liver, gallbladder and right kidney during abdomen scans helps level the playing field for users across experience levels
				

AI in the department – quality of imaging

Fetal Diagnosis

Measure fetal biometry parameter in one click

BiometryAssist™ 1, a feature based on Deep Learning technology, is an automatic technology for biometric measurement. It enables users to measure the fetal growth parameters with one click while maintaining exam consistency.



An automated reporting tool for fetal heart diagnosis

HeartAssist™ 1, a feature based on Deep Learning technology, provides automatic classification of ultrasound image into measurement views required for heart diagnosis and provides measurement results.



An automated classification of the images and annotation of the structures

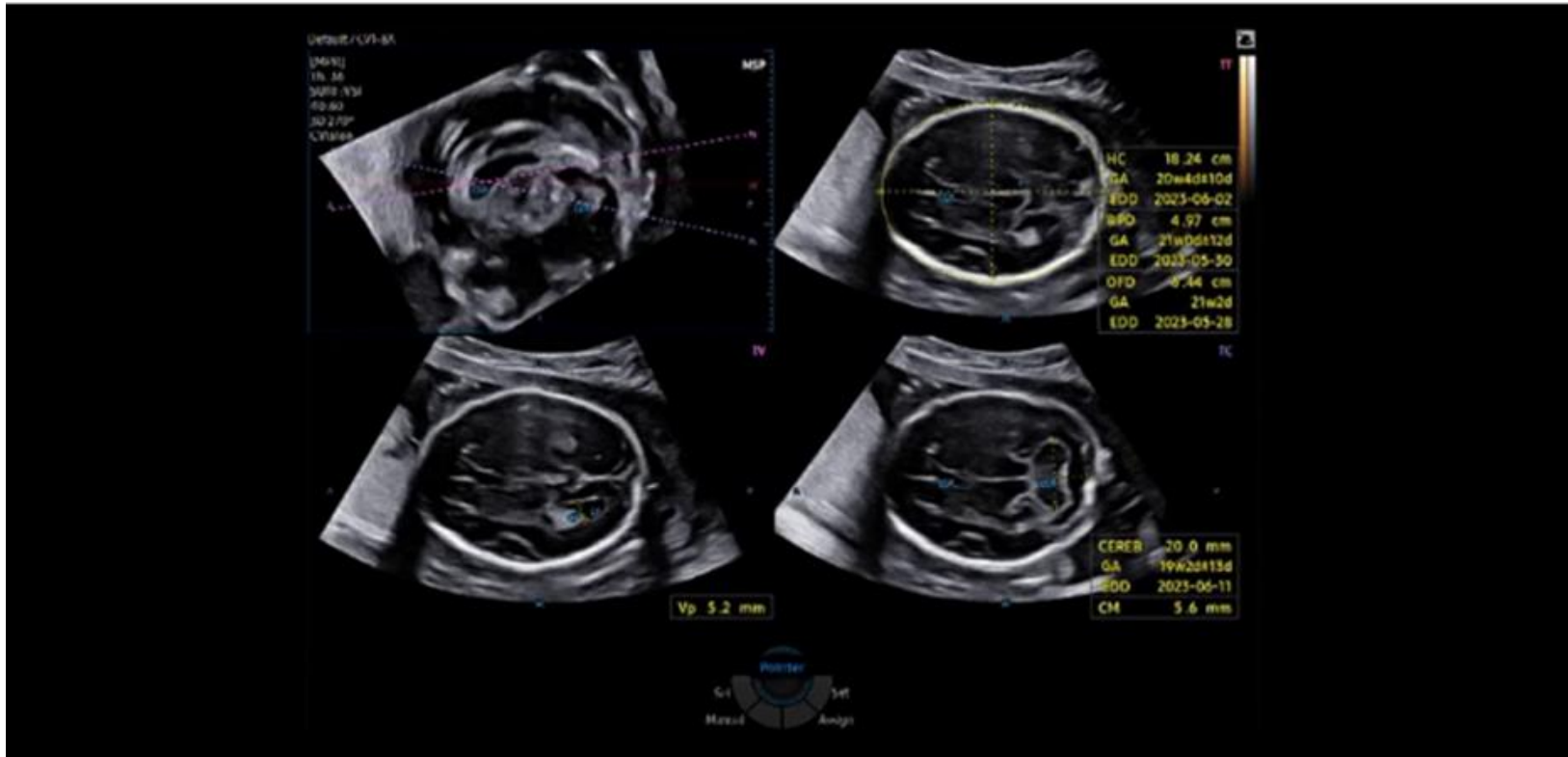
ViewAssist™ 1 a feature based on Deep Learning technology, provides automatic classification of the ultrasound images and annotation of the structures to help healthcare professionals in convenient measurement.



AI in the department – ease of measuring



Measure fetal brain in one click



AI at the business meeting – workflow & finance

Staff shortages are plaguing hospitals and clinics worldwide. To add to the challenge, the demand for ultrasound services is on the rise. Clinicians and administrators need a solution, and they need it now.

27,000

additional sonographers will be needed in the US by 2024, an increase of 24%ⁱ



By 2030,

it is anticipated that there will be a shortfall of 10 million healthcare workersⁱⁱ



81%

of health systems surveyed in the U.S. reported radiology technologist shortagesⁱⁱⁱ



90%

of sonographers experienced work-related musculoskeletal disorders^{iv}



AI can improve both the patient and clinician experience

Reduce exam time by 81%

and help properly align and display recommended views and measurements of the fetal brain^v



Complete a wide range of Doppler measurements with up to

93% fewer keystrokes

by using Cardiac Auto Doppler with AI Spectrum Recognition^{ix}



AI is the teammate that clinicians need to make informed decisions faster, and that administrators need to meet increased demand, and support ultrasound users' well-being.

How did we get here?

- Historically, much interest centred on fetal anomaly detection, particularly brain & cardiac abnormalities, rather than fetal measurements
- Most algorithms use retrospective rather than prospective datasets
- Research data typically based on single centre. Decision making for algorithm made by experienced expert(s). Applicable in 'routine' environments?
- Studies of 3rd trimester size, growth & EFW poorly represented
- Published data:
 - Fetal brain abnormalities
 - Improving detection of severe cardiac abnormalities
 - Fetal biometry
- Consider:
 - Bias in decision making during research & development & therefore its potential impact on
 - commercially available systems

How did we get here?

2nd trimester brain & cardiac abnormalities

Use of real-time artificial intelligence in detection of abnormal image patterns in standard sonographic reference planes in screening for fetal intracranial malformations

M. LIN^{1#}✉, X. HE^{2#}, H. GUO^{3#}, M. HE¹, L. ZHANG¹, J. XIAN⁴, T. LEI¹✉, Q. XU³, J. ZHENG¹, J. FENG¹, C. HAO⁵, Y. YANG¹, N. WANG⁶✉ and H. XIE¹

- Excellent performance achieved in identifying 10 types of intracranial image pattern
- Performance comparable to expert sonologists in diagnosis
- System's assessment required significantly less time (0.025s) compared to experts (4.4s) to read an image

9 different patterns of abnormality as identified using ISUOG reference planes & normal :

1. *normal*
2. *no cavum septum pellucidum*
3. *no septum pellucidum*
4. *crescent shaped single ventricle*
5. *mild ventriculomegaly*
6. *severe ventriculomegaly*
7. *mega cisterna magna*
8. *open 4th ventricle*
9. *intraventricular cyst*
10. *non-intraventricular cyst*

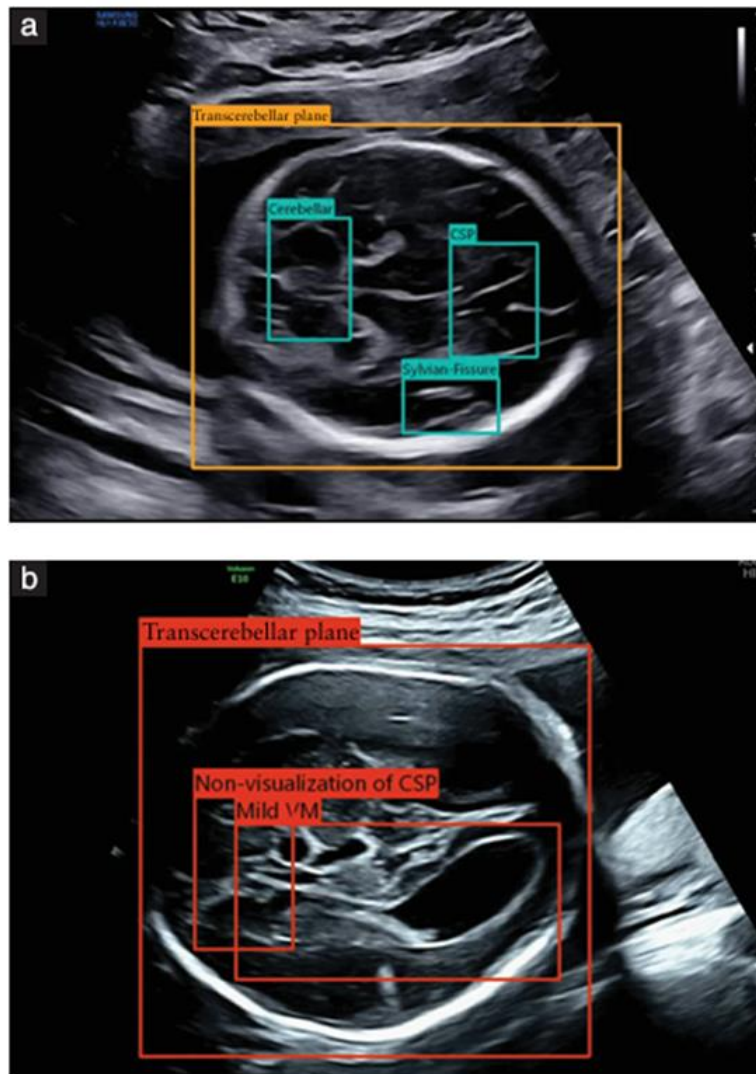


Figure 1 Two representative standard planes of the fetal head, extracted from ultrasound video, highlighting structures and pathology classified automatically by the real-time artificial intelligence system developed by Lin *et al.*². (a) Transcerebellar plane, showing automatic identification of the cerebellum, Sylvian fissure and cavum septi pellucidi (CSP) in a normal fetus. (b) Trans-ventricular plane in a fetus with mild ventriculomegaly (VM) and non-visualization of the CSP.

Multicentre, retrospective study – 3 centres, China

- Normal fetuses & with CNS abnormalities
- 18-40 wks
- Images Jan 2010 – Dec 2018, videos Jan – Sep 2020
- 9 different abnormal patterns & normal in reference planes
- Development & validation of AI system - 43,890 images from 16,297 pregnancies, 169 videos from 166 pregnancies
- Training images assessed for quality control & labelling:
 - a) 5 sonologists, 5-10 yrs experience (>5000 fetal exams each)
 - b) 2 senior, >10yrs (>10,000 fetal US exams each)
- Image preparation 7 months – selection 3/12, labelling 4/12
- 4hrs /day, not at w/e
- a) grader 1 min/image = 718hrs for 43,078 images, 2/12 each
- b) 10s to check = 120hrs, 5days/month then 10days/month
- 3 days to feed DL computer continuously
- 1 day for fine tuning

System validation

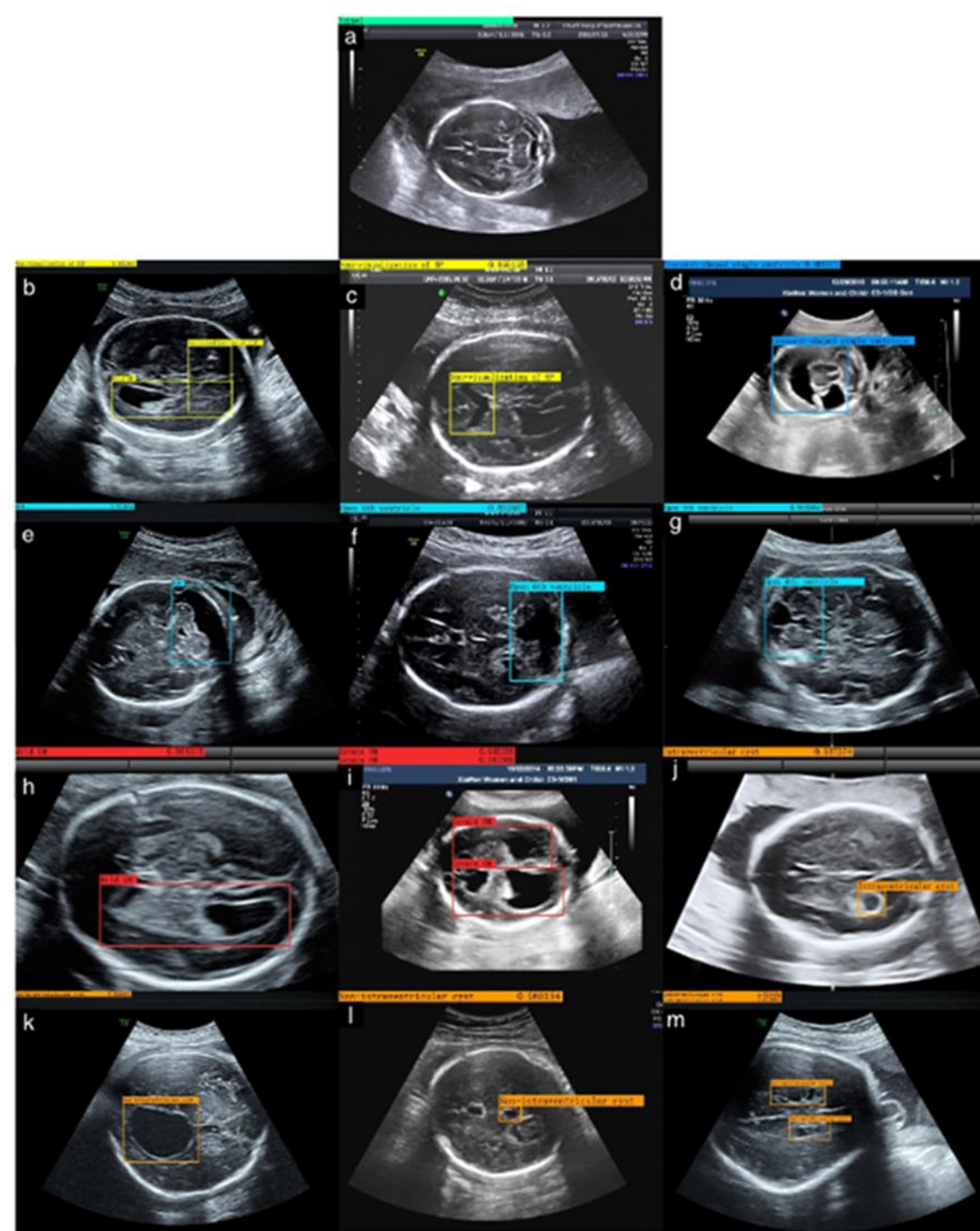
- External validation used videos as prospective data
- 13 sonologists each viewed same dataset
- All blinded to image diagnoses
- Varying experience
 - 4 experts – professors, >10yrs, >10,000 FA scans
 - 4 competent
 - 5 trainees - residents, 2-4 years, >1000 FA scans

Results - system on par with experts overall

- better than experts for mild & severe VM (offline measuring tool inconvenient?)
- slightly better than competent
- superior to trainees

Conclusions

- Direct connection to US machine allows analysis in R/T
- Has potential to improve CNS detection rates, particularly where operators have little formal training



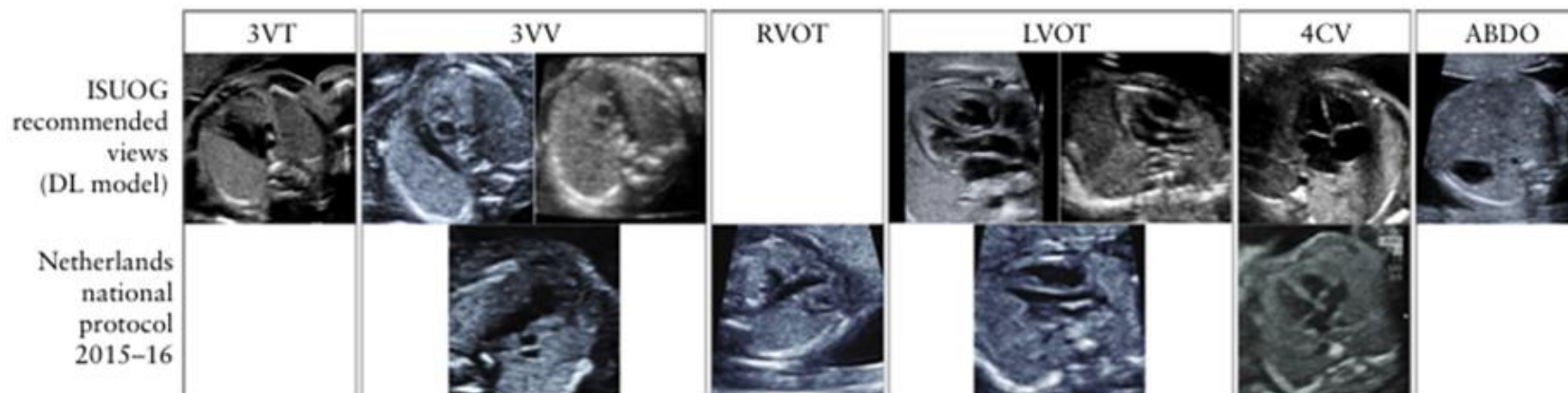
Deep-learning model for prenatal congenital heart disease screening generalizes to community setting and outperforms clinical detection

2024

C. ATHALYE¹, A. VAN NISSELROOIJ², S. RIZVI¹, M. C. HAAK² and R. ARNAOUT^{1,3,4}

- NW Netherlands 2015-2016
- routine anomaly scan includes 4 cardiac view images
- cardiac images of 108 cases reviewed:
- heart normal = 42
- CHD = 66
 - 35 identified
 - 31 missed

- non-blinded grading - 2 experts
 - completeness & technical correctness (0-5)
 - defect visible yes/no
 - duration, image number etc
 - 1 expert labelled each image frame
- blinded diagnosis - 3 experts (15-25yrs)
 - normal v CHD
- DL model
 - had been already trained for previous study
 - 47 lesion types had already encountered



Deep-learning model for prenatal congenital heart disease screening generalizes to community setting and outperforms clinical detection

C. ATHALYE¹, A. VAN NISSELROOIJ², S. RIZVI¹, M. C. HAAK²®, A. J. MOON-GRADY³ and R. ARNAOUT^{1,3,4}®

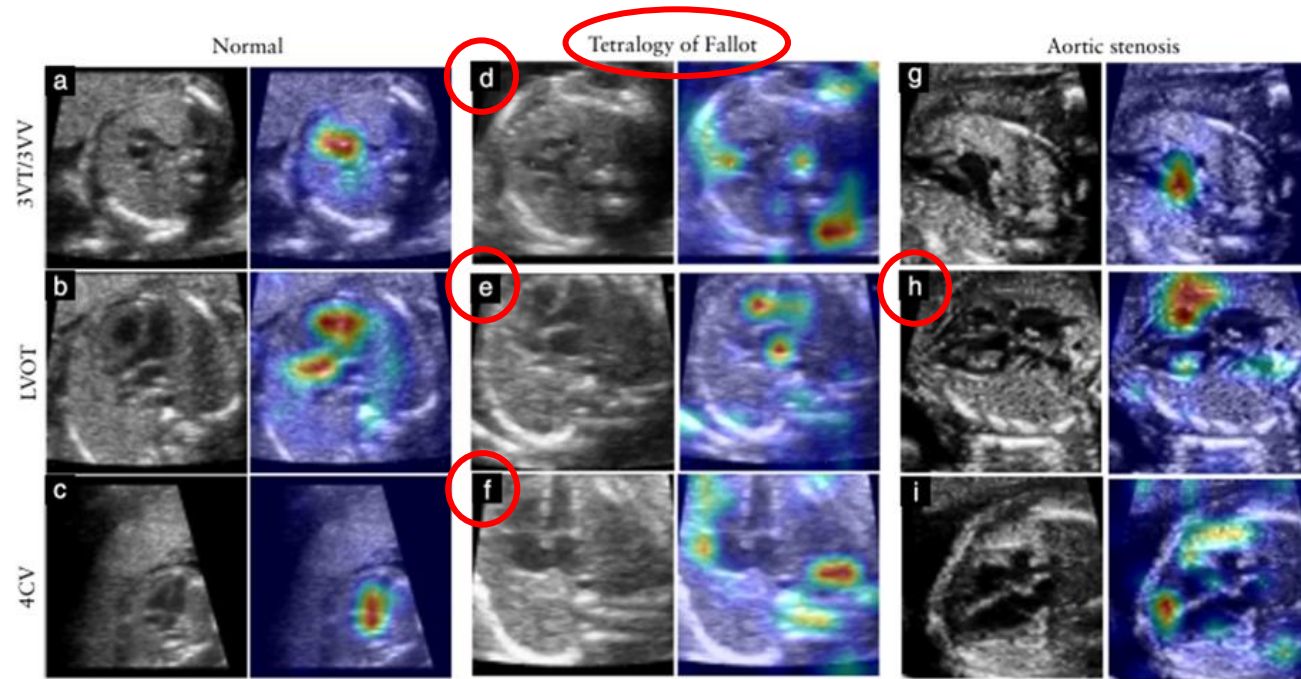


Figure 2 Diagnostic classifier in studies that the model got correct and blinded experts missed. Model-labeled views (grayscale) with corresponding GradCAM images representing heat maps showing areas of the image most important in model decision-making (red shows most important areas). The model correctly identified normal views (a–c), focusing on the aorta (a), left ventricular outflow tract (LVOT) and right ventricle (b), and interatrial and interventricular septa (c). The model identified the abnormal three-vessel-and-trachea (3VT) view and LVOT (d,e) and abnormal four-chamber view (4CV) cardiac axis (f) in tetralogy of Fallot, and abnormal LVOT in aortic stenosis (h). The human experts misclassified these tetralogy and aortic stenosis patients as 'normal'. 3VV, three-vessel view.

- blinded diagnosis - 3 experts (15-25yrs)
 - normal v CHD
 - identified 32/42 normal hearts
 - diagnosed 25/35 abnormal hearts
 - detected 11/31 missed diagnoses
- DL model
 - 47 lesion types had already encountered
 - detected 16 of 19 not previously encountered
 - identified 32/42 normal hearts
 - diagnosed 32/35 abnormal hearts
 - detected 27/31 missed diagnoses

Deep-learning model for prenatal congenital heart disease screening generalizes to community setting and outperforms clinical detection

C. ATHALYE¹, A. VAN NISSELROOIJ², S. RIZVI¹, M. C. HAAK²✉, A. J. MOON-GRADY³ and R. ARNAOUT^{1,3,4}✉

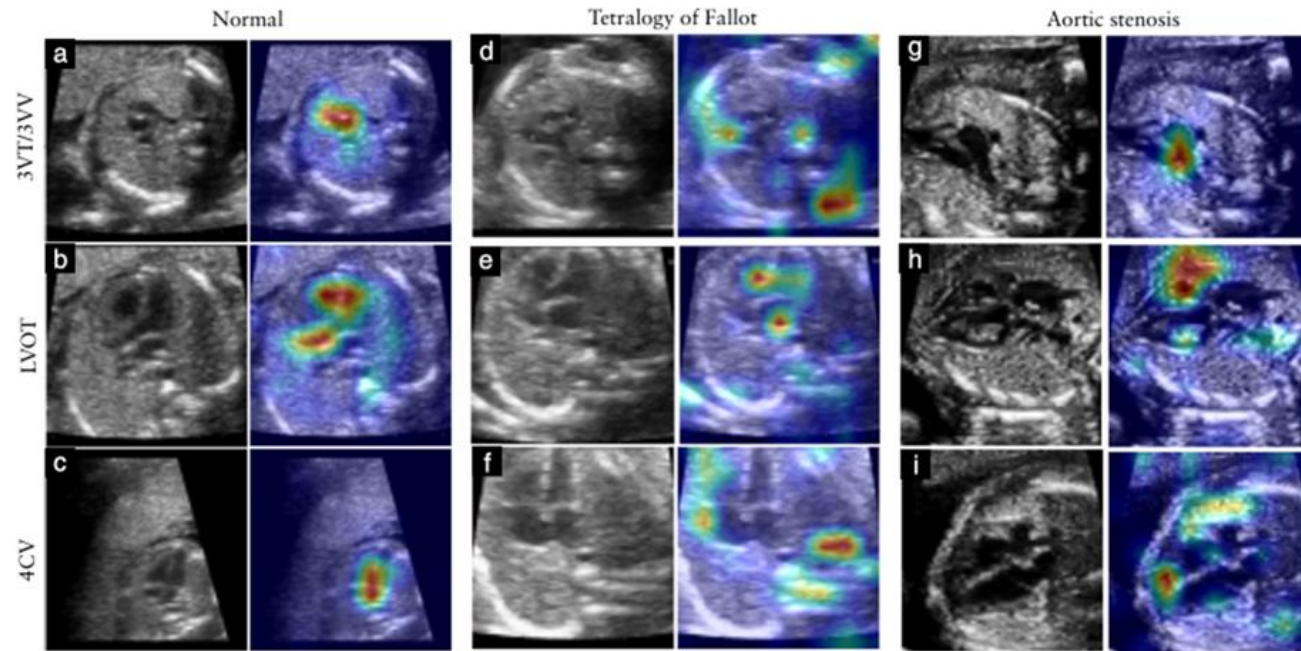


Figure 2 Diagnostic classifier in studies that the model got correct and blinded experts missed. Model-labeled views (grayscale) with corresponding GradCAM images representing heat maps showing areas of the image most important in model decision-making (red shows most important areas). The model correctly identified normal views (a–c), focusing on the aorta (a), left ventricular outflow tract (LVOT) and right ventricle (b), and interatrial and interventricular septa (c). The model identified the abnormal three-vessel-and-trachea (3VT) view and LVOT (d,e) and abnormal four-chamber view (4CV) cardiac axis (f) in tetralogy of Fallot, and abnormal LVOT in aortic stenosis (h). The human experts misclassified these tetralogy and aortic stenosis patients as 'normal'. 3VV, three-vessel view.

Conclusions:

1. A previously trained DL algorithm outperformed routine cardiac screening in cohort where >50% CHD initially missed
2. AI model outperformed human experts in blinded trial
3. DL models have potential to improve CHD detection in routine screening of low risk populations

How did we get here?

Automated measurements

Role of artificial-intelligence-assisted automated cardiac biometrics in prenatal screening for coarctation of aorta

C. A. TAKSØE-VESTER^{1,2,3}, K. MIKOLAJ⁴, O. B. B. PETERSEN^{1,2}, N. G. VEJLSTRUP⁵,
A. N. CHRISTENSEN⁴, A. FERAGEN⁴, M. NIELSEN⁶, M. B. S. SVENDSEN³ and
M. G. TOLSGAARD^{1,2,3} ©

- 60% cases of isolated CoA not identified antenatally
- Cardiac images from 4 regions of Denmark, 2008-2018
- 18-22 wks

- AI model developed to:
 - identify standard cardiac planes
 - perform automated cardiac biometric imaging
 - RV, LV
 - diameters PA & Aao
 - RV/LV ratio, MPA/Aao ratio

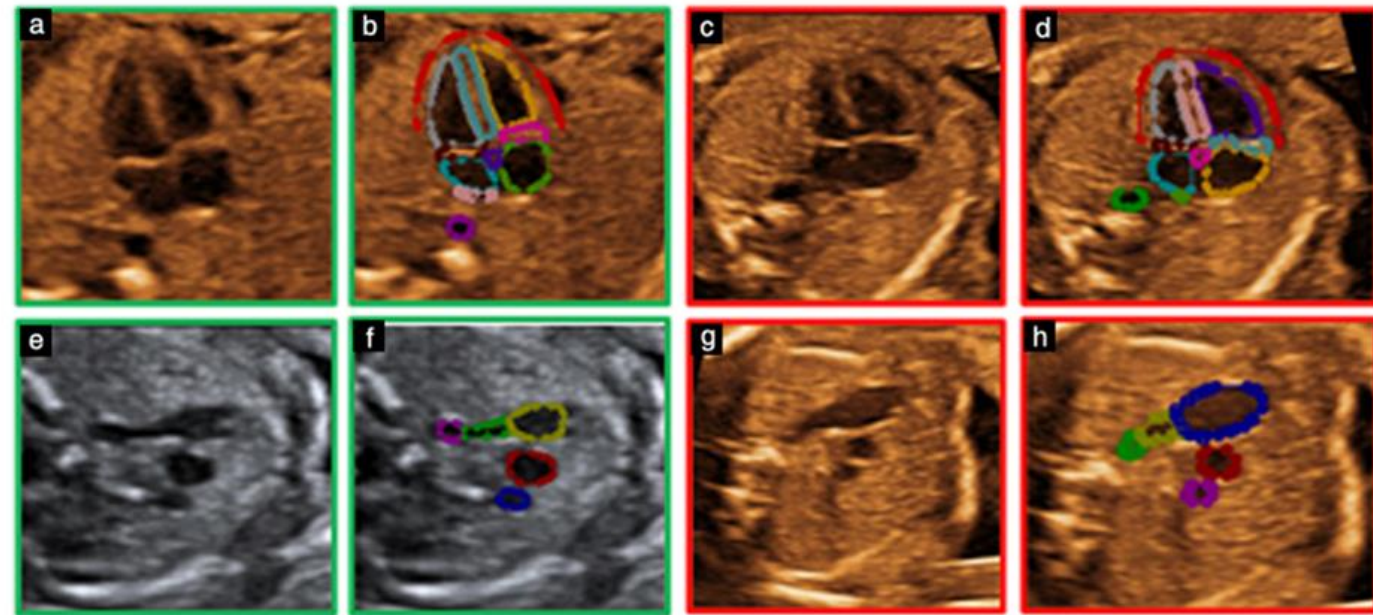


Figure 1 Four-chamber view (a–d) and three-vessel view (e–h) ultrasound images: examples of artificial intelligence segmentation (b,d,f,h) on which automated measurements are based, in a healthy control fetus (a,b,e,f) and a fetus with coarctation of the aorta (c,d,g,h).

- 73 cases of CoA paired with health controls, ratio 1:100

Role of artificial-intelligence-assisted automated cardiac biometrics in prenatal screening for coarctation of aorta

C. A. TAKSØE-VESTER^{1,2,3}, K. MIKOLAJ⁴, O. B. B. PETERSEN^{1,2}, N. G. VEJLSTRUP⁵,
A. N. CHRISTENSEN⁴, A. FERAGEN⁴, M. NIELSEN⁶, M. B. S. SVENDSEN³ and
M. G. TOLSGAARD^{1,2,3} ©

Comparing CoA & normal biometrics

- Aao & Dao significantly smaller in CoA
- RV & RV/LV ratio significantly larger in CoA
- MPA/Aao ratio significantly larger in CoA

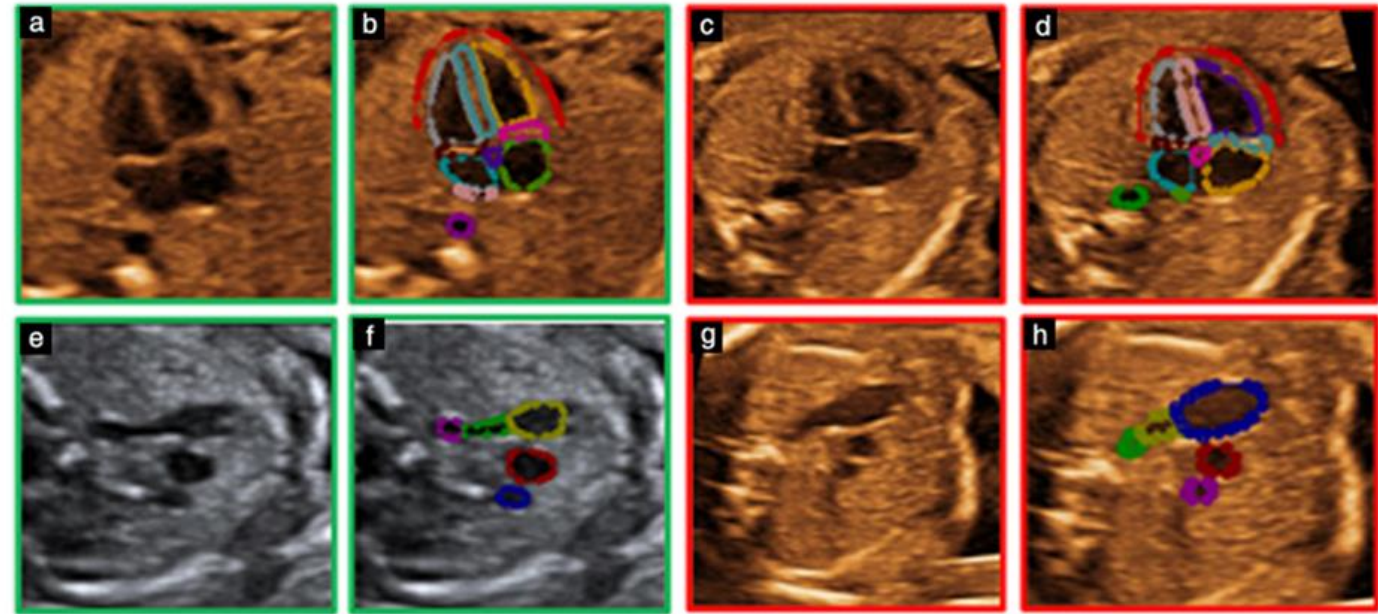


Figure 1 Four-chamber view (a–d) and three-vessel view (e–h) ultrasound images: examples of artificial intelligence segmentation (b,d,f,h) on which automated measurements are based, in a healthy control fetus (a,b,e,f) and a fetus with coarctation of the aorta (c,d,g,h).

Conclusion:

Intregation of AI automated cardiac biometric measurements has potential to enhance substantially the performance of screening for, & detection of, fetal CoA at 18-22 wks

How did we get here?

Routine anomaly scan - FASP standard planes

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

Jacqueline Matthew^{1,2} | Emily Skelton^{1,2,3} | Thomas G. Day^{1,2} |
 Veronika A. Zimmer¹ | Alberto Gomez¹ | Gavin Wheeler¹ | Nicolas Toussaint¹ |
 Tianrui Liu⁴ | Samuel Budd⁴ | Karen Lloyd¹ | Robert Wright¹ | Shujie Deng¹ |
 Nooshin Ghavami¹ | Matthew Sinclair¹ | Qingjie Meng⁴ | Bernhard Kainz⁴ |
 Julia A. Schnabel¹ | Daniel Rueckert^{4,5} | Reza Razavi^{1,2} | John Simpson^{1,2} |
 Jo Hajnal¹

- Prospective study, London UK
- May 2019 – March 2020, previous normal FA scan
- 23 women, each scanned twice – AI assisted & standard scans
- 2 experienced operators (20yrs combined)
- 13 FASP standard planes, BPD, HC, AC, FL
- AI:
 - assisted automated image acquisition, biometry & reporting
 - traffic light system provided additional information for each anatomical view
 - provided 5 candidate images for each plane
 - sonographer selected best quality plane, automatic biometry included
 - autoreport generated

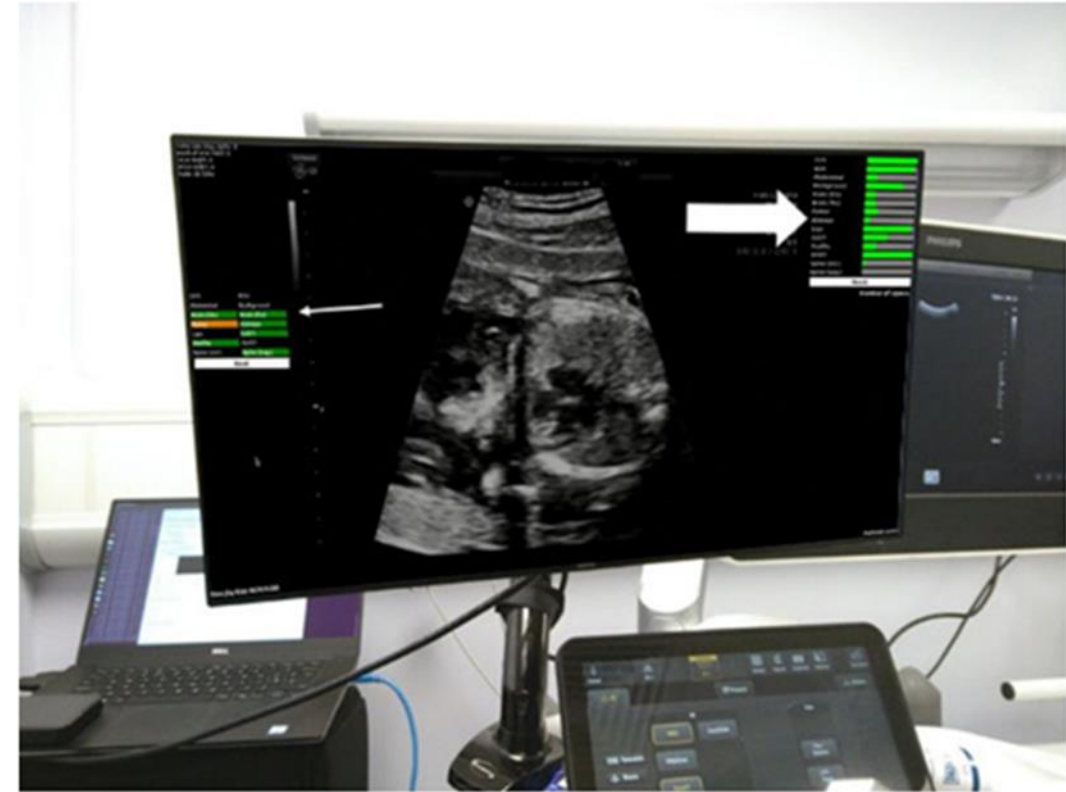


FIGURE 2 Research clinic room set up and display monitors. Large white arrow: AI feedback overlay, displaying real-time detection confidence for each standard view. Small white arrow: 'Traffic light' system indicating the overall confidence of the completeness of the data capture for each standard view (high, moderate, low)

- 299 manual & 260 AI-assisted FASP standard images
- AI-assisted scan times 34.7% shorter than manual
- 14.32 v 21.93mins , mean saving time of 7.62mins
- AI tools easy to use
- Change in scanning approach perceived
- AI tools made it easier to concentrate on pt interaction during the scan
- AI assisted report included 93% of 4 FASP biometry views, 98% in manual report
- AI successfully saved 73% of required images, 98% in manual report



Conclusions:

- Separating freehand scanning from image capture & measurement -> faster scan time & altered workflow
- Removing repetitive tasks may allow more attention to be directed to identifying fetal abnormalities
- Further work required to improve image plane detection algorithm

How did we get here?

Fetal biometry

Evaluation of automated tool for two-dimensional fetal biometry

2019

I. SALIM¹✉, A. CAVALLARO¹✉, C. CIOFOLO-VEIT², L. ROUET², C. RAYNAUD², B. MORY², A. COLLET BILLON², G. HARRISON³, D. ROUNDHILL³ and A. T. PAPAGEORGHIU¹✉

¹Nuffield Department of Obstetrics and Gynaecology, University of Oxford, Oxford, UK; ²Philips Research, Paris, France; ³Philips Ultrasound, Bothell, WA, USA

- Jan – Dec 2015, Oxford UK
- 99 pregnancies 20-40wks
- 3 HC (transthalamic), AC & FL images per case
- Automatic measurement system developed
- Caliper placement:
 - acceptable 53.3%
 - minor adjustment needed 35.6%
 - major adjustment required 10.1%
- No caliper adjustment needed:
 - HC 73.1%
 - FL 71.0%
 - AC 15.8%
- Failed to recognise & measure AC in nine images
- No failures for HC & FL

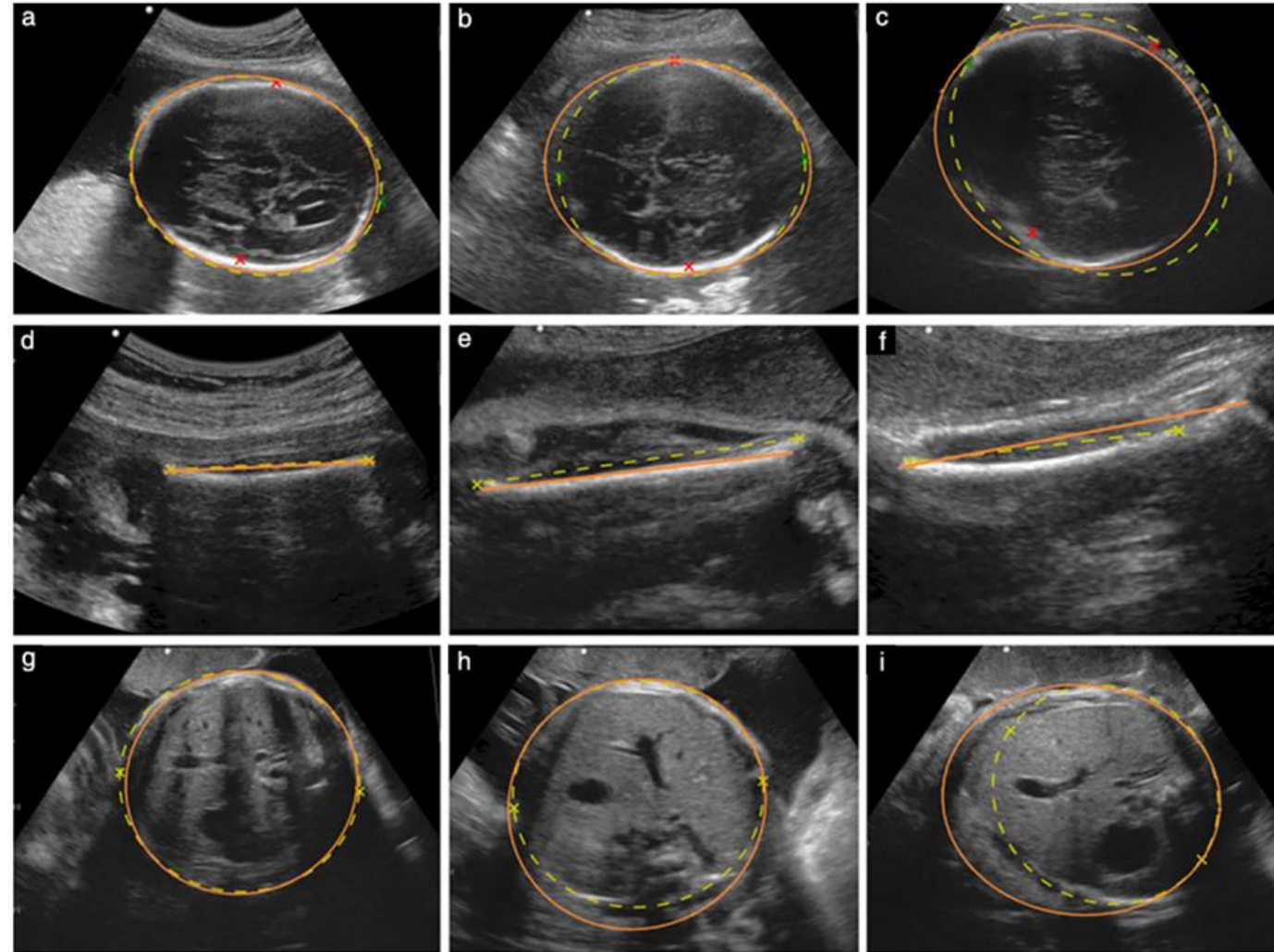


Figure 1 Manual (—) and automated (---) contour delineation for measurement of fetal head circumference (HC) (a–c), femur length (FL) (d–f) and abdominal circumference (AC) (g–i). Good fit between manual and automated calipers is observed in (a), (d) and (g). In (b), (e) and (h), minor adjustment of automated calipers is required. In (c), (f) and (i), major adjustment of automated calipers is required for accurate measurement.

I. SALIM¹ , A. CAVALLARO¹ , C. CIOFOLO-VEIT², L. ROUET², C. RAYNAUD², B. MORY², A. COLLET BILLON², G. HARRISON³, D. ROUNDHILL³ and A. T. PAPAGEORGHIU¹ 

¹Nuffield Department of Obstetrics and Gynaecology, University of Oxford, Oxford, UK; ²Philips Research, Paris, France; ³Philips Ultrasound, Bothell, WA, USA

Conclusion:

1. Automated tool correctly identified biometric variable in 99% of images
2. Resulting AI measurements high degree of accuracy (HC -0.81%, AC 2.40%, FL 3.76%)
3. Measurements exhibited bias, with tool underestimating biometry
4. This could be overcome by further algorithm improvements
5. Adjustable calipers for manual correction remains a requirement

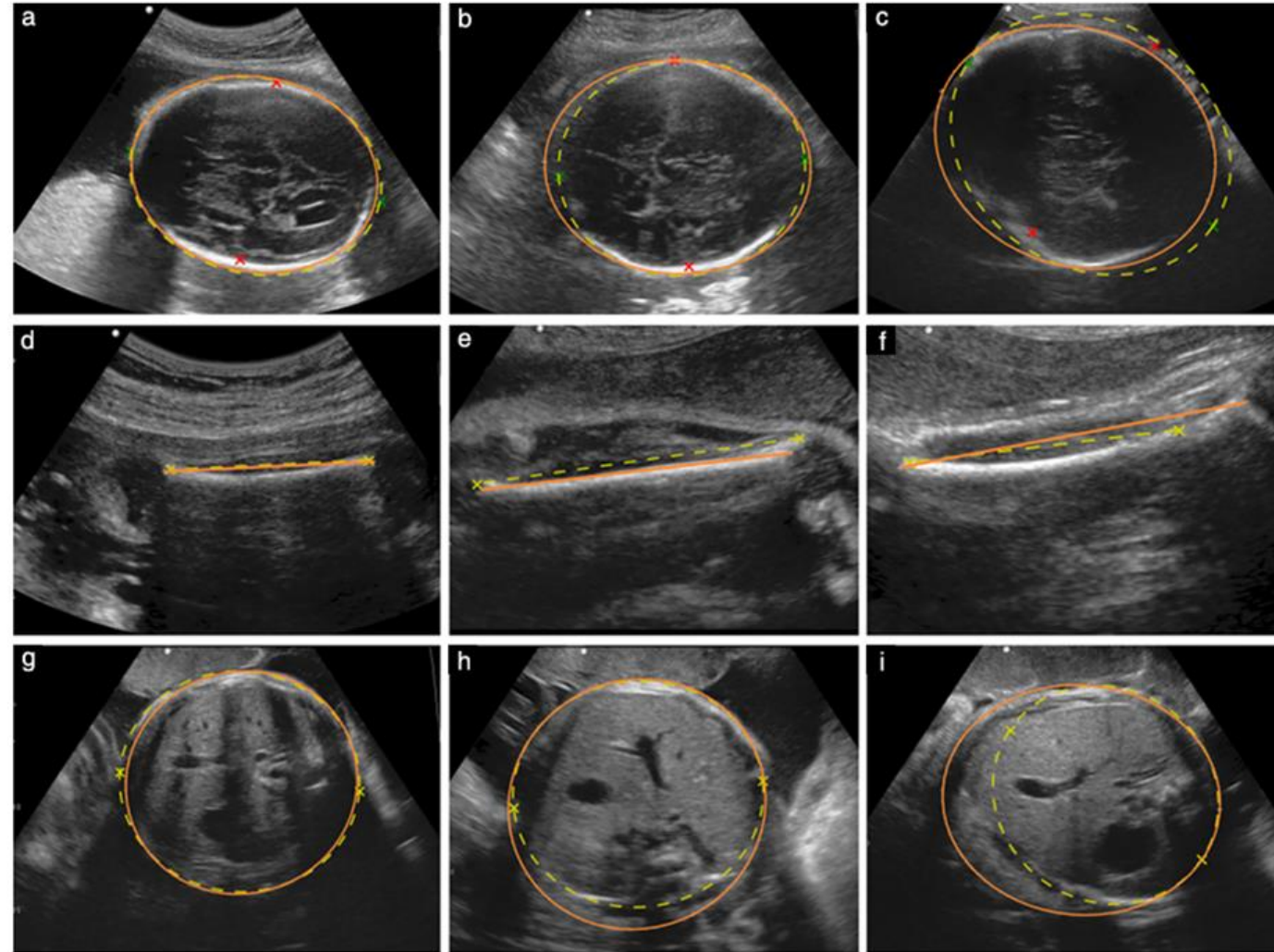


Figure 1 Manual (—) and automated (---) contour delineation for measurement of fetal head circumference (HC) (a–c), femur length (FL) (d–f) and abdominal circumference (AC) (g–i). Good fit between manual and automated calipers is observed in (a), (d) and (g). In (b), (e) and (h), minor adjustment of automated calipers is required. In (c), (f) and (i), major adjustment of automated calipers is required for accurate measurement.

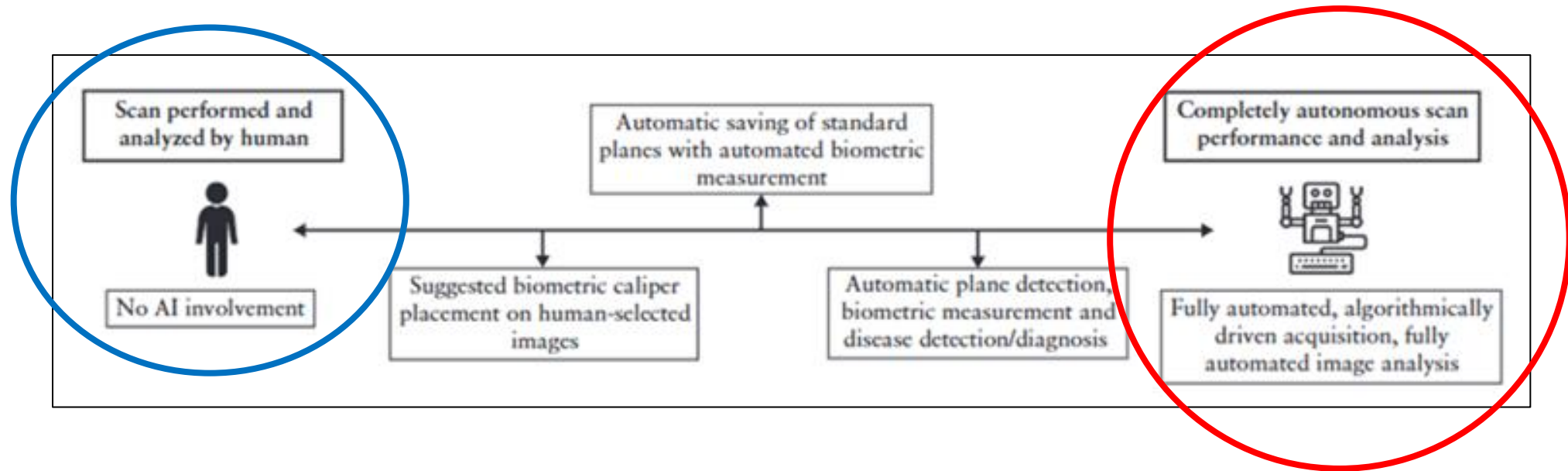


AI is the teammate that clinicians need to make informed decisions faster, and that administrators need to meet increased demand, and support ultrasound users' well-being.

Where are we now?

- AI imaging - shows great potential in improving anomaly detection but bias inevitable
 - evaluation from prospective & low risk data needed
 - systems can & will make mistakes
- AI biometry – routine biometry showing potential but bias inevitable,
 - AC data least good
 - systems can & will make mistakes
- Training – AI options already available
 - who is making the above decisions for commercially available systems?
 - how much variation is there in the decisions made by different systems?
- AI assisted workflow – exam time reduced, supports change in operator activities
- What are the implications for gaining technical skills & US relevant knowledge for current sonographers & for future sonographers?

Where could we be going?

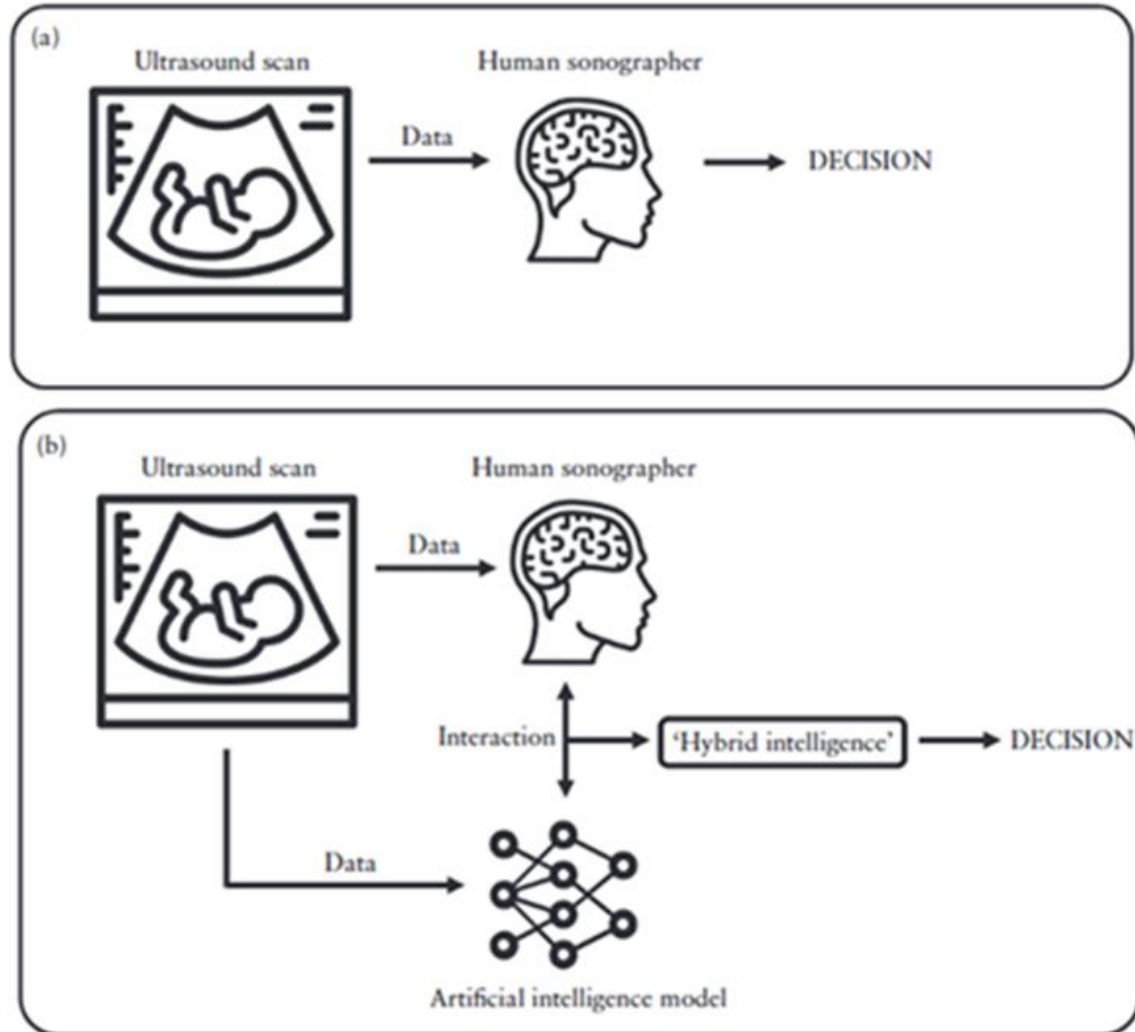


Opinion

Sonographer interaction with artificial intelligence: collaboration or conflict?

T. G. DAY^{1,2}*, J. MATTHEW²*, S. BUDD²,
J. V. HAJNAL², J. M. SIMPSON^{1,2}, R. RAZAVI^{1,2}
and B. KAINZ^{2,3,4}

Where should we be going?

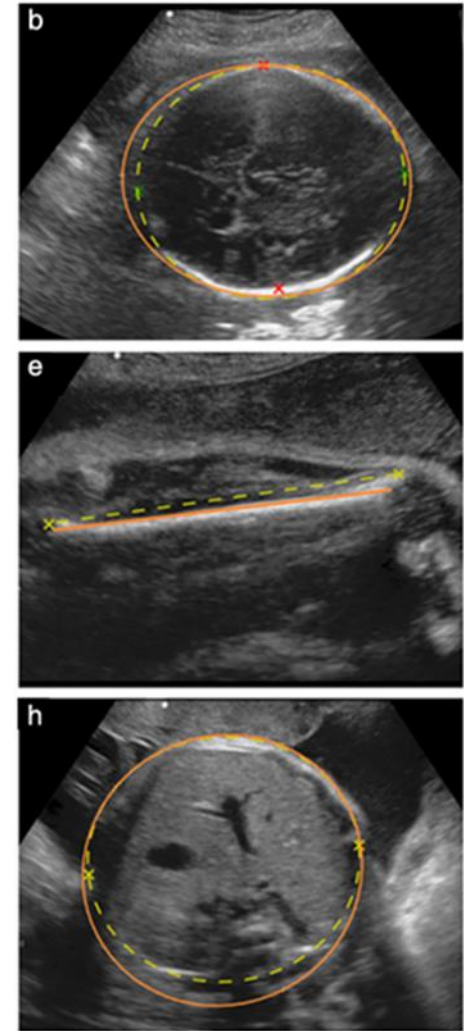
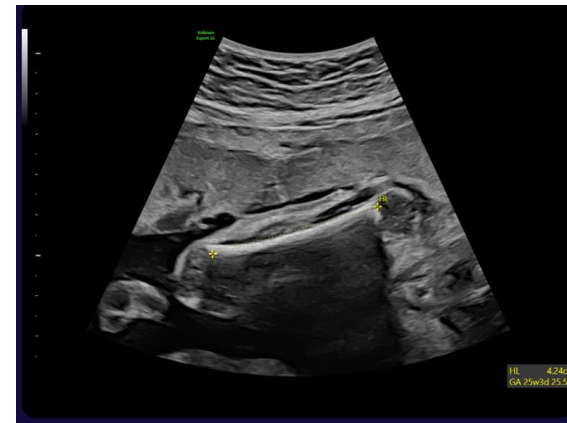
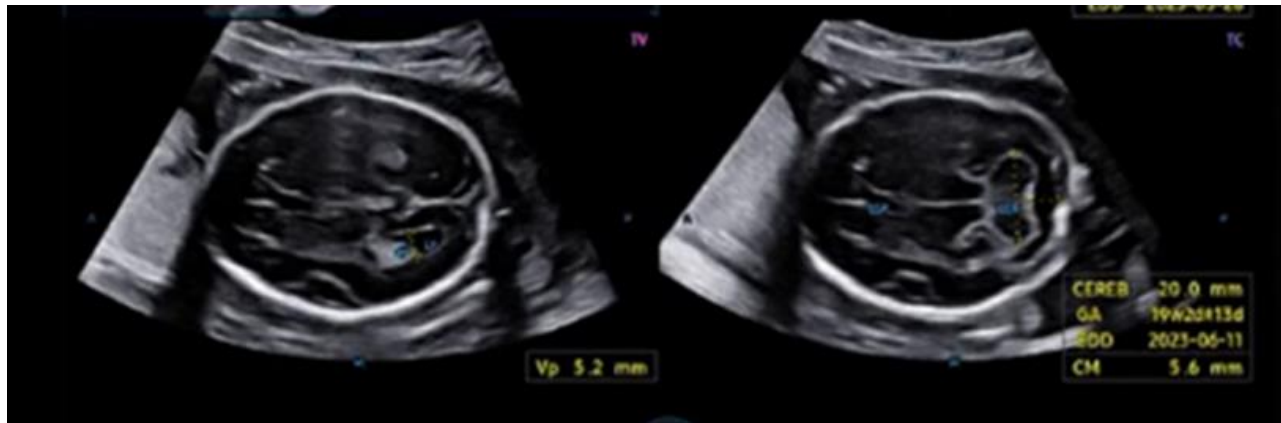


Opinion

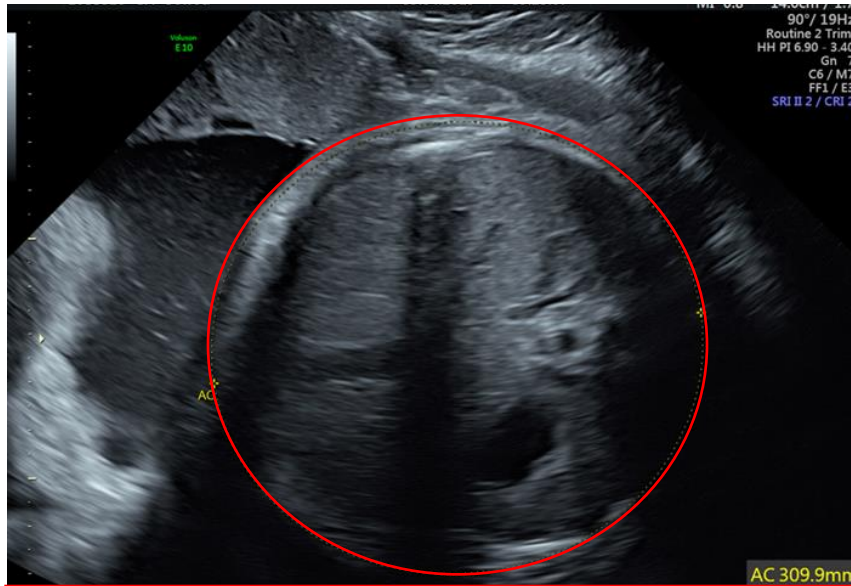
When we've got there

How will AI-assisted operators:

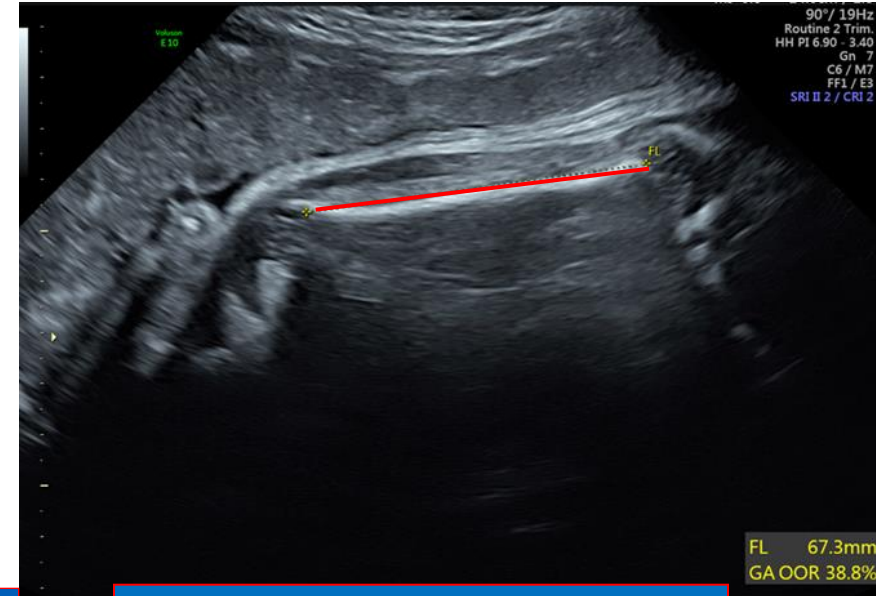
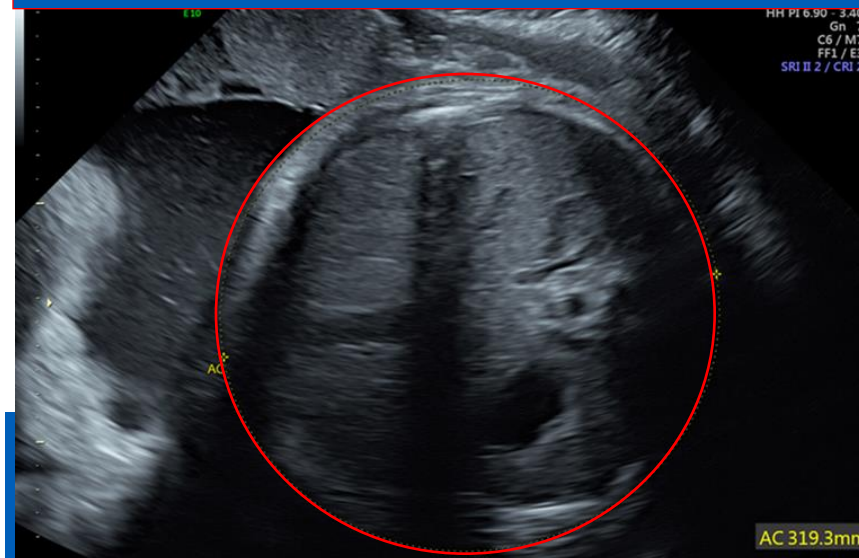
- learn how to obtain the correct planes?
- learn how to take measurements correctly?
- know when the anatomy doesn't look normal?
- learn how to distinguish between poor technique & a genuine problem?
- what will they be able to teach their students?



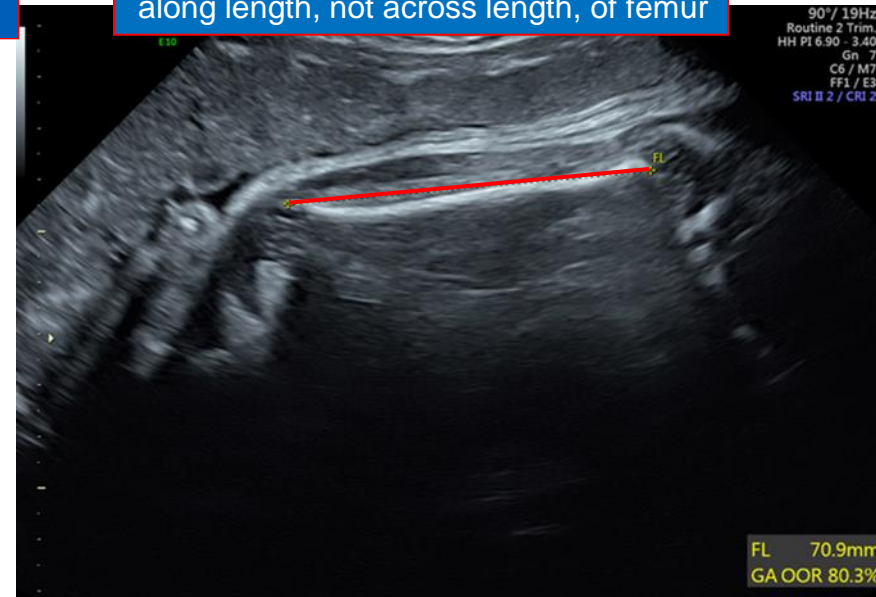
Caliper placement for growth velocity & EFW



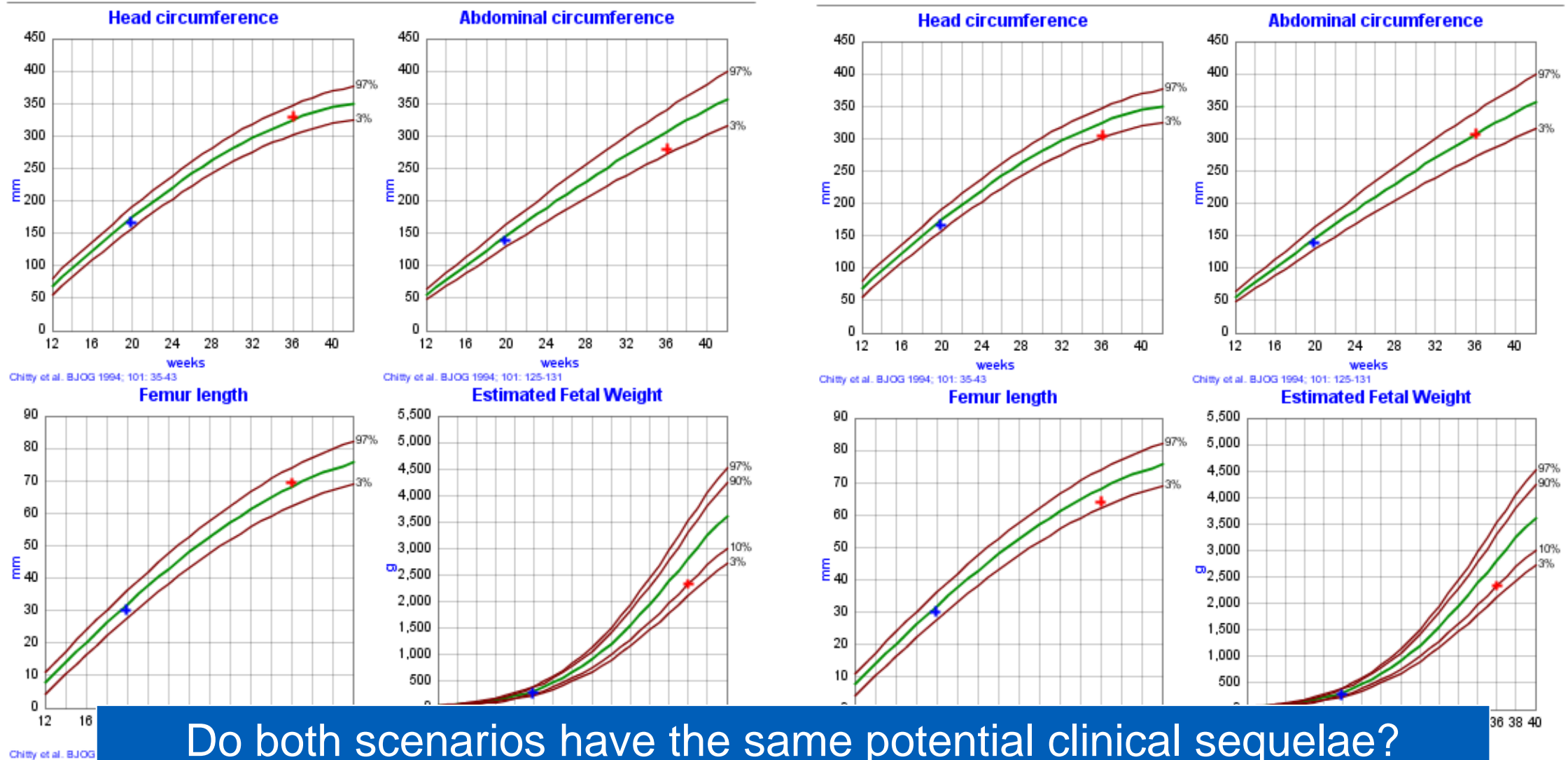
deciding which echoes represents skin line posteriorly of AC



along length, not across length, of femur

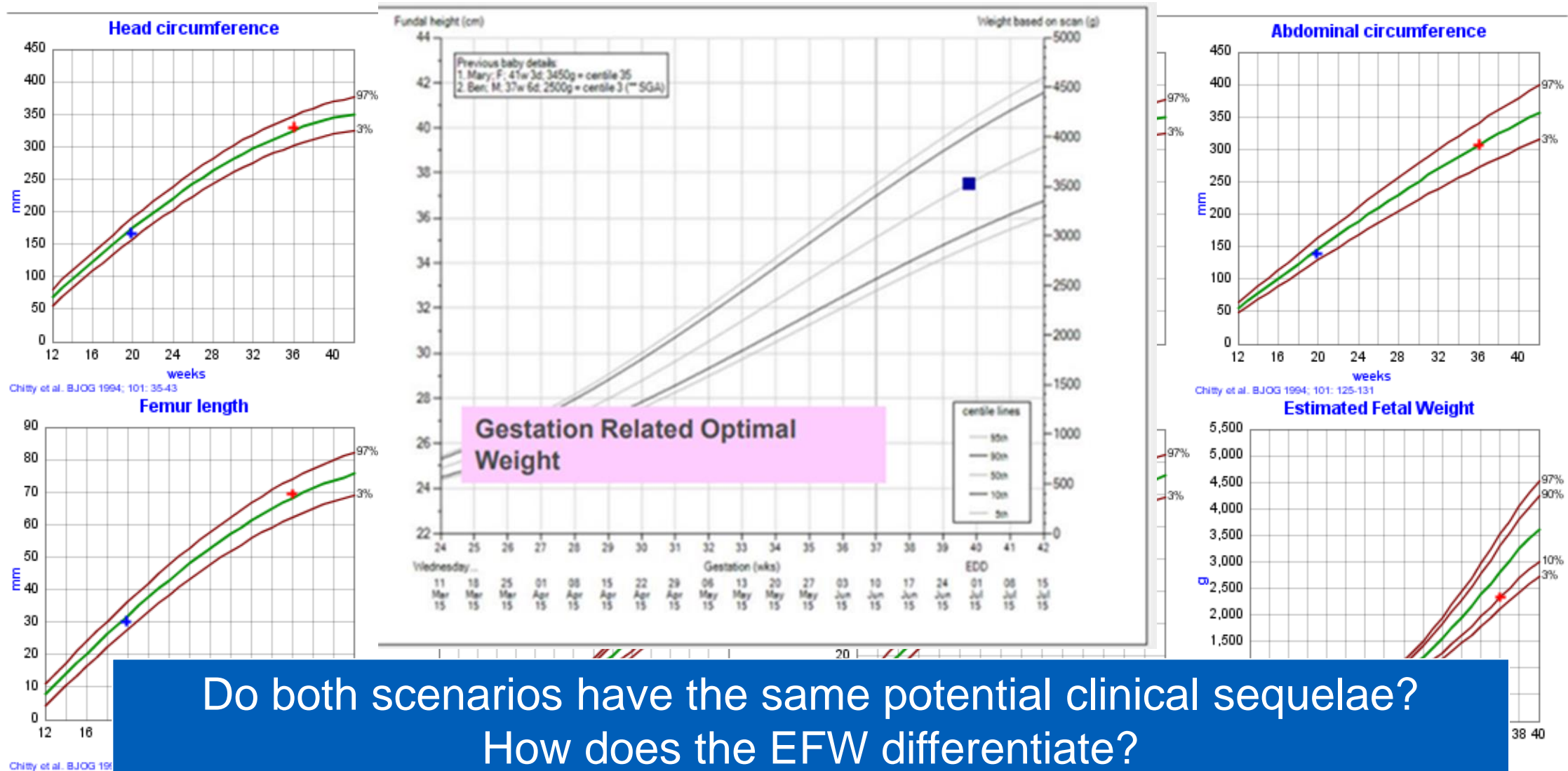


HC, AC, FL & EFW - or EFW alone for clinical management?



Do both scenarios have the same potential clinical sequelae?
How does the EFW differentiate?

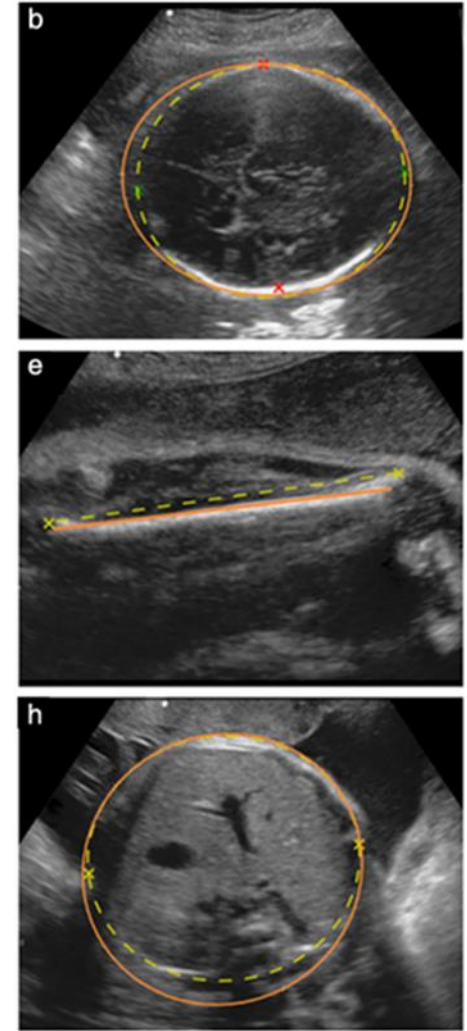
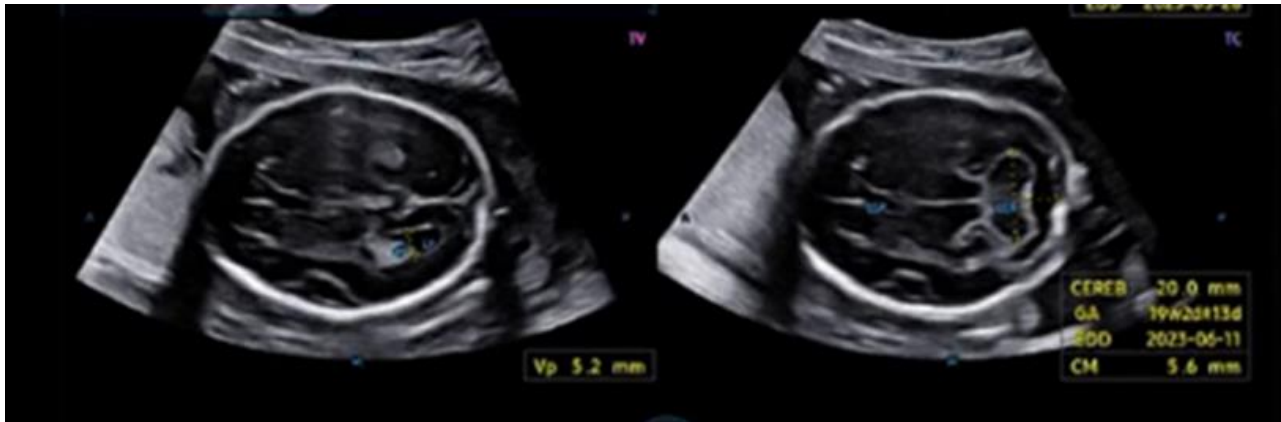
HC, AC, FL & EFW - or EFW alone for clinical management?



Do both scenarios have the same potential clinical sequelae?
How does the EFW differentiate?
How does GROW differentiate?

When we've got there

- What has the effect of GAP/GROW been on understanding of parameters contributing to EFW?
- Does this matter – to the pregnancy, to the clinicians, to you?
- What will the impact of AI-assisted measurements & imaging decisions be on what future sonographers will learn, know & be able to apply?
- Skills rarely used are of doubtful value. What happens when they are needed in a critical situation?



Who is making the decisions?

- Who are teaching & checking the AI programmes during their development?
- Do they all have the appropriate expertise?
- Do they all agree?
- Whose decisions go forward for implementation & why?
- Is there a potential mismatch between commercial & clinical priorities?
- Is there international/national/professional body guidance?
- Does this guidance agree?
- Who or what is responsible when things go wrong?

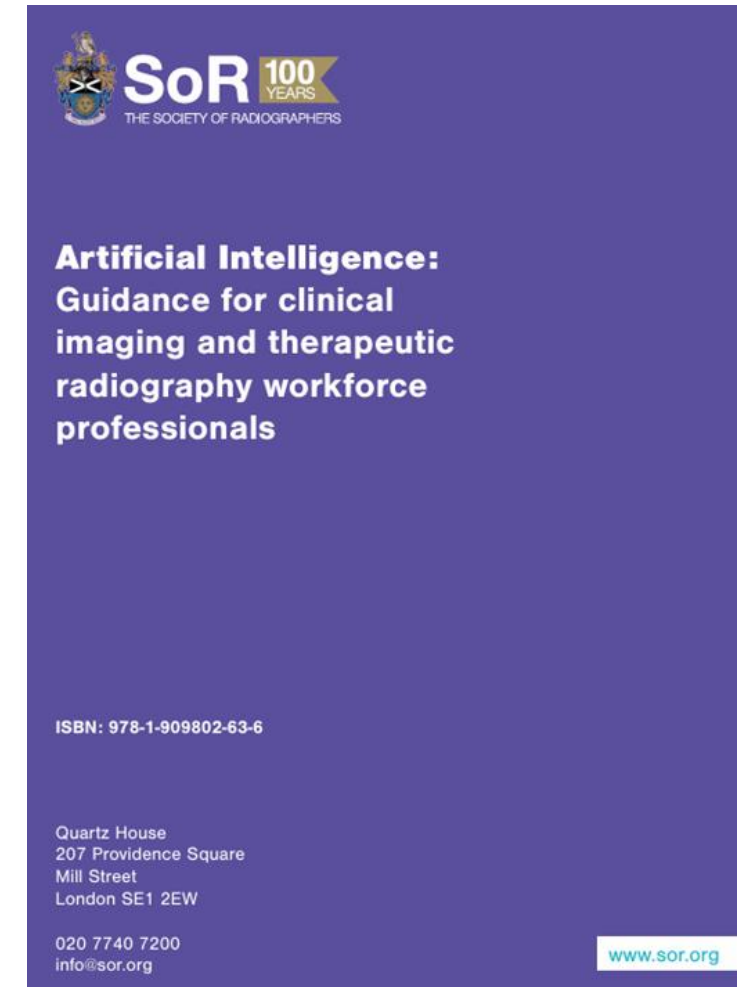
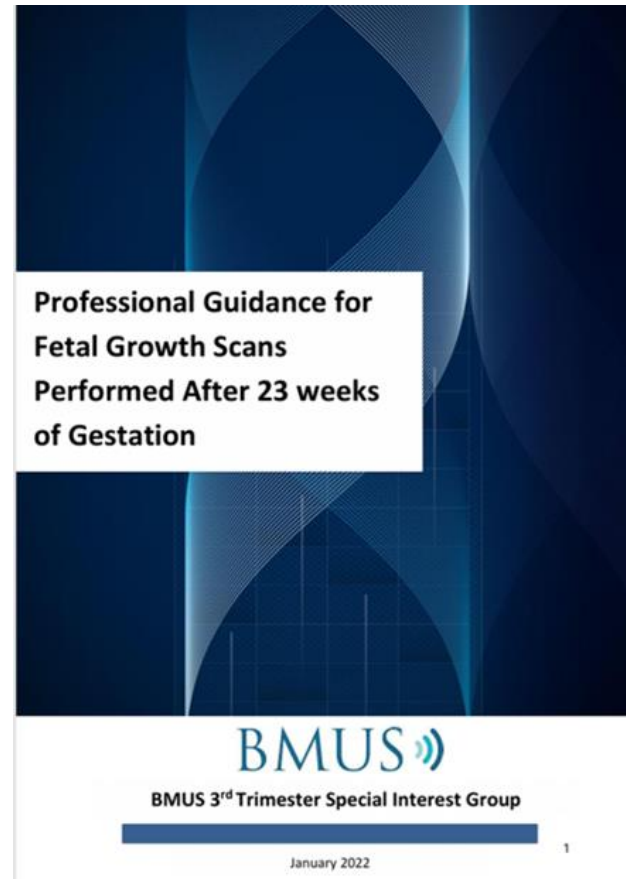
Who is setting the standards?

ISUOG - no statement to date
(August 2024)

BMUS – no published guidance but
SIG3 group addressing issue in
revision of ‘growth guidelines’

AXREM – to be tabled at Nov 2024
meeting

SCoR – yes, 2021



Who is setting the standards?

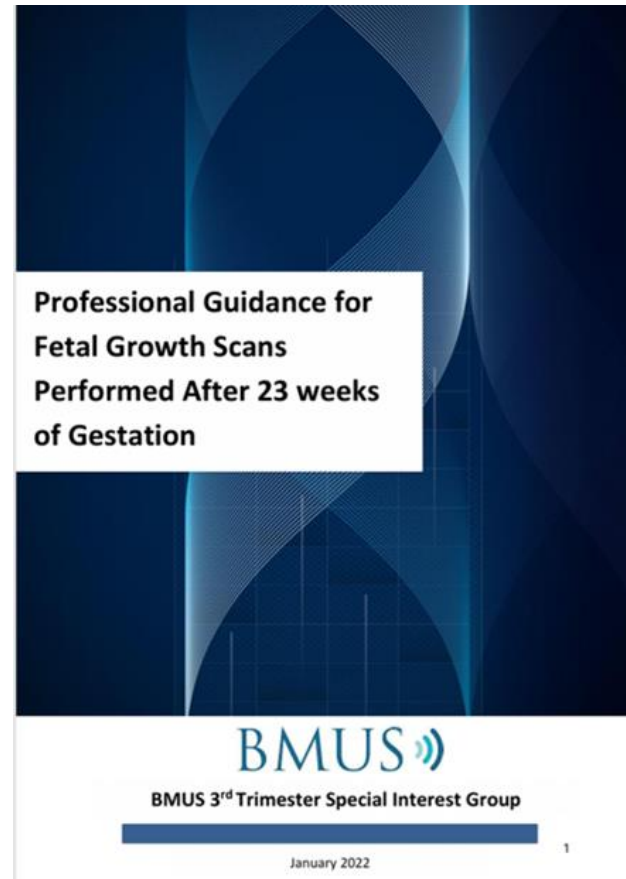
ISUOG - no statement to date (August 2024)

BMUS – no published guidance but SIG3 group addressing issue in revision of 'growth guidelines'

AXREM – to be tabled at Nov 2024 meeting

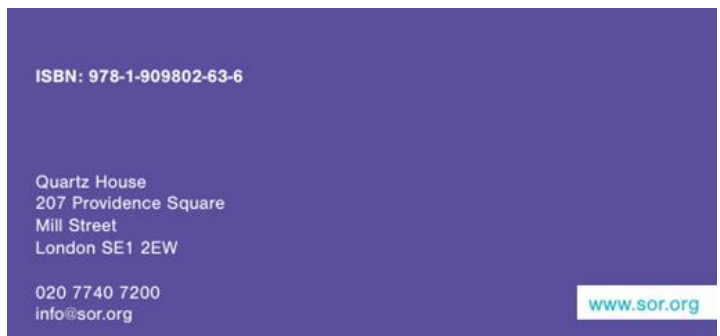
SCoR – yes, 2021

Establishment of international, clinically based standards needed urgently – before it's too late

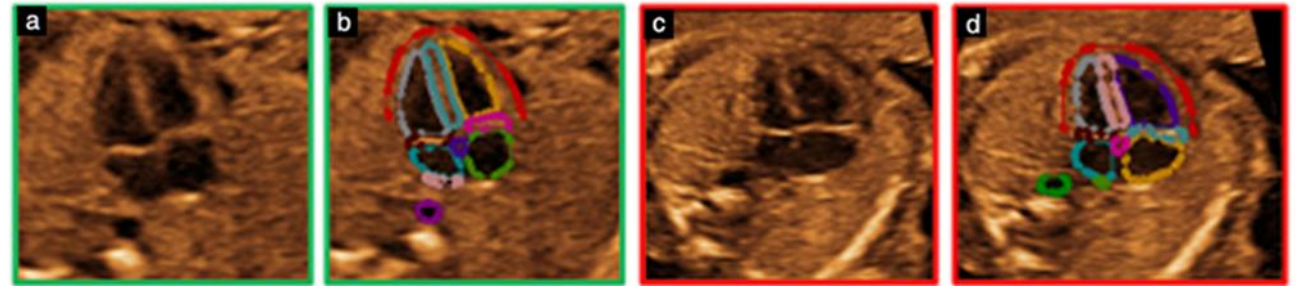


All AI systems that are deployed clinically require human oversight of their implementation

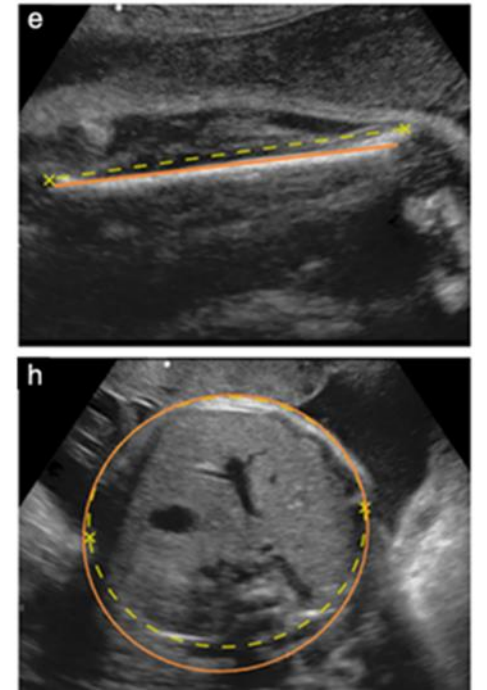
Radiography staff must learn to evaluate, interact & oversee the actions of AI driven tools within their workflow (p 19)



In conclusion



- Who is/are making the decisions for AI systems?
- What standards are being applied? Are they right?
- Where should responsibility lie for missed or incorrect diagnoses due to AI?
- Will the standard of obstetric imaging drift up or down with time as the technical expertise of the workforce changes - especially when AI trained trainers start training the next generation?
- Embracing AI provides many opportunities & challenges – are you ready to participate in making sure it works for our routine ultrasound service?



Reduce exam time by 81%
and help properly align and
display recommended views and
measurements of the fetal brain*



Thank you for your attention