

REVIEW

Artificial intelligence, fetal echocardiography, and congenital heart disease

Thomas G. Day^{1,2}  | Bernhard Kainz³ | Jo Hajnal¹ | Reza Razavi^{1,2} | John M. Simpson^{1,2}

¹Faculty of Life Sciences and Medicine, School of Biomedical Engineering and Imaging Sciences, King's College London, London, UK

²Department of Congenital Cardiology, Evelina London Children's Healthcare, Guy's and St Thomas' NHS Foundation Trust, London, UK

³Department of Computing, Faculty of Engineering, Imperial College London, London, UK

Correspondence

Thomas G. Day, School of Biomedical Engineering and Imaging Sciences, King's College London, North Wing, St Thomas' Hospital, Westminster Bridge Road, London, SE1 7EH, UK.

Email: Thomas.day@kcl.ac.uk

Funding information

Wellcome Trust / EPSRC Centre for Medical Engineering, Grant/Award Number: WT 203148/Z/16/Z; Wellcome Trust, Grant/Award Number: 102431

Abstract

There has been a recent explosion in the use of artificial intelligence (AI), which is now part of our everyday lives. Uptake in medicine has been more limited, although in several fields there have been encouraging results showing excellent performance when AI is used to assist in a well-defined medical task. Most of this work has been performed using retrospective data, and there have been few clinical trials published using prospective data. This review focuses on the potential uses of AI in the field of fetal cardiology. Ultrasound of the fetal heart is highly specific and sensitive in experienced hands, but despite this there is significant room for improvement in the rates of prenatal diagnosis of congenital heart disease in most countries. AI may be one way of improving this. Other potential applications in fetal cardiology include the provision of more accurate prognoses for individuals, and automatic quantification of various metrics including cardiac function. However, there are also ethical and governance concerns. These will need to be overcome before AI can be widely accepted in mainstream use. It is likely that a familiarity of the uses, and pitfalls, of AI will soon be mandatory for many healthcare professionals working in fetal cardiology.

Key Points

What's already known about this topic?

- Artificial intelligence (AI) is rapidly becoming part of everyday life, and interest is growing in its use in improving medical outcomes.
- AI may be a useful tool in fetal cardiology to improve rates of antenatal diagnosis of congenital heart disease, as well as other potential benefits.
- Potential pitfalls exist however, and future clinicians will likely need to have a thorough understanding of the risks and benefits of AI.

What does this study add?

- This review article summarizes the fundamentals of AI, the potential uses of AI in fetal cardiology, and what the future may hold.

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.

1 | ULTRASOUND TO SCREEN FOR CONGENITAL HEART DISEASE: A VITAL TOOL, BUT STILL FAILING

The use of ultrasound to image the fetal heart was first reported in 1964, initially using M-mode techniques to characterize fetal heart rate and heart size.¹ The routine use of B-mode ultrasound to accurately diagnose structural congenital heart disease (CHD) in the fetus began in the 1980s, with groups in the UK and USA publishing case series demonstrating the utility of this technique.^{2,3} Since then, there have been constant incremental technical developments with the introduction of spectral Doppler, color Doppler, and three-dimensional (3D) imaging techniques, all of which are now in widespread clinical use. Fetal echocardiography in expert hands is a highly sensitive and specific diagnostic test.^{4,5} However, when views of the fetal heart are incorporated as part of anomaly screening programs, both sensitivity and specificity in the detection of CHD are lowered substantially.

Fetal echocardiography is now considered a core component of the routine fetal anomaly scan. Although obstetric practice varies widely, most countries worldwide offer a mid-trimester fetal anomaly screening ultrasound scan with the aim of detecting serious malformations, and international guidelines recommend that such scans include specific views of the fetal heart.^{6,7} Despite this, antenatal CHD detection rates remain lower than for most other major structural anomalies. International registry-based data suggest a wide variation in antenatal detection rates, with some countries detecting only 14% of severe CHD cases before birth.⁸ Significant geographical variation within countries has also been demonstrated.^{9,10}

This is a problem, because evidence suggests that infants with many forms of serious CHD diagnosed postnatally rather than antenatally are less likely to survive long enough to undergo heart surgery, are less likely to survive after such surgery, and are more likely to have an adverse long-term neurological outcome.¹¹⁻¹⁵ In addition, accurate antenatal diagnosis allows the parents to make an informed decision regarding the continuation of pregnancy, and can also allow therapeutic intervention in selected cases.¹⁶

There are complex reasons behind the failure of fetal echocardiography to achieve universal antenatal detection of CHD. Recent data have shown that the most frequent reasons for CHD to be overlooked during routine mid-trimester anomaly scans are poor adaptational skills of the sonographer to acquire and optimize the correct sonographic plane, or failure to recognize an abnormality which is present on the ultrasound image.¹⁷ Thus, in the majority of cases of missed CHD, either the correct cardiac view was not correctly obtained, or the defect was clearly demonstrated but not recognized by the operator.¹⁷ Previous work has shown the positive impact of operator experience and programs of staff training to improve recognition of cardiac lesions.¹⁷⁻²¹ However, such programs are labor and time intensive and need to be repeated with staff turnover.

A complementary strategy would be to alter the paradigm, for example by making ultrasound systems “smarter” via the integration

of artificial intelligence (AI). Such an approach could have the potential to assist the sonographer in recognizing cardiac abnormalities, whilst remaining unobtrusive, quick, and easy to learn. By combining this approach with continued education and training, it may be possible improve antenatal detection rates of CHD through multiple mechanisms simultaneously.

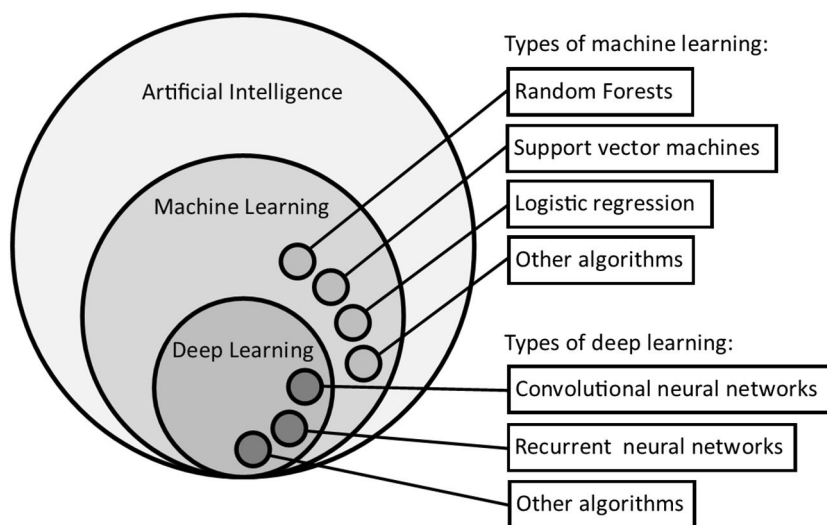
2 | CORE CONCEPTS IN AI

AI can be defined broadly as the field of science that aims to use computer programs to learn complex tasks and make predictions based on data.²² Although the field of AI research has existed for over 70 years, in the last decade there has been an “AI boom” with extremely rapid progress in multiple fields.²³ This has been largely driven by three factors: (1) hardware development, with the production of affordable graphical processing units, optimized to perform a huge amount of simultaneous calculations, (2) the growing collection and availability of “big data” (especially data that has been labeled), essential in order to train AI systems, and (3) the application of complex AI methods, including neural networks, which we will discuss below.

Machine learning (ML) can be seen as an integral part of AI. ML can be defined as the use of computer programs that automatically improve with experience, so over time will become more successful in their defined task. A more formal way of stating this is to say a machine can be said to learn from experience E with respect to a task T and performance measure P , if its performance of T , as measured by P , improves with E .²⁴ Deep learning (DL) is a specific type of ML that uses neural networks (explained below) arranged into many layers (typically more than five, up to hundreds).²⁵ Each layer can extract more abstract and high-level features from the input data, allowing complex interpretation and prediction from the supplied data, for example image classification in the field of computer vision. Figure 1 outlines the relationship between AI, ML, and DL, showing some examples of ML methods such as random forests, support vector machines, and logistic regression, and examples of DL methods such as convolutional neural networks and recurrent neural networks. Many other ML methods have been developed, all which have strengths and weaknesses when applied to a specific problem. A full description of these methods is beyond the scope of this review, but we refer the reader to excellent online learning resources.²⁶

ML methods can be broadly categorized into “supervised” and “unsupervised” learning methods. Supervised learning is best used when large amounts of labeled training data are available, and the algorithm uses this data to learn how to make specific predictions when presented with new, unlabeled data. “Unsupervised learning” is a different approach, where the ML algorithm is shown unlabeled data and asked to identify clusters within it, without generating specific predictions. This review is mostly concerned with supervised learning, as this predominates in the medical AI literature, however unsupervised learning can be a powerful tool in the identification of previously unseen patterns within patient data.

FIGURE 1 The relationship between artificial intelligence, machine learning, and deep learning, with examples of machine learning and deep learning methods



Neural networks represent the current state-of-the-art in the field of medical AI, and the development of these models has resulted in super-human performance in certain medical tasks (although whether this translates to an actual clinical improvement is debatable, as we will discuss later).²⁷ There are many varieties of neural networks that have been developed to perform optimally in specific tasks (e.g., convolutional networks for computer vision, recurrent networks for language processing). A full discussion of these network architectures is also outside the scope of this review, but more information can be found at the online learning resources mentioned above.²⁶

Neural networks use labeled data (X , the raw data; Y , the label) to develop a complex model that describes the relationship between X and Y . The basic building block of the neural network is the perceptron, first described in 1958. A perceptron takes the weighted sum of multiple inputs (X), using a vector of weights (W), adds a bias (b), and passes the resulting sum through an activation function, to give an output.²⁵ These weights and biases are known as the *parameters* of the network, and these are what the neural network is learning. Neural networks are composed of multiple perceptrons, arranged into multiple layers, with deeper layers of the network receiving the outputs of the previous layer as an input. Figure 2 demonstrates how imaging data from a fetal echocardiogram might be incorporated into a neural network, and how the neural networks are composed of multiple layers of perceptrons.

Neural networks are trained starting with random values for the parameters. Training involves presenting the network with training data (e.g. image data) and comparing the network predictions to the known ground truth, to calculate an error. Using this error, each parameter can be altered slightly in order to reduce the error. As more labeled training data is passed through, over time the performance of the network will improve in an iterative fashion. The final performance of the network can then be tested by passing previously unseen data (test data) through it and measuring how the predictions compare to the known ground truth values.

3 | AI IN MODERN MEDICINE

Early attempts to use AI in medicine were disappointing, with rule-based systems unable to cope with the complexity of medical scenarios, and performed poorly when confronted by large volumes of new data.^{22,25} The development of neural networks has resulted in more success, although real-world implementations of AI solutions are scarce, with the majority of research papers testing the performance of algorithms using retrospectively acquired labeled data.²⁸ In very well-defined, consistent, and repetitive tasks (such as X-ray or retinal/skin photograph classification) ML algorithms have been shown to achieve or even exceed human expert-level performance.²⁹⁻³³

Adult cardiology is another area that has shown rapid development in the use of AI, raising hope that fetal cardiology may have similar promise.^{34,35} Both specialties have a large focus on echocardiography, which has the potential to utilize AI to give real-time feedback to the clinician. Models have been developed that can classify the correct echocardiographic view, and also identify specific pathology within these views.^{36,37}

The automatic quantification of cardiac function (in postnatal cardiology) using AI is another area of cardiology that has received much interest, both as a potential to reduce the inter- and intra-observer variability seen in current practice, and to reduce the time taken to perform the study.^{35,38} Commercially available solutions that utilize AI are now available.^{39,40} These AI-based models can sit unobtrusively “on-cart” within the ultrasound machine, aiding the sonographer by automatically measuring 3D cardiac chamber volumes and ejection fraction. Such methods have been shown to be reproducible, and by reducing or removing human error, provide a means to help standardization across different clinicians or institutions.³⁹ There is also the potential to alter and streamline echocardiography workflow, as automated measurements are faster, and allow the sonographer to focus on other aspects of the study.

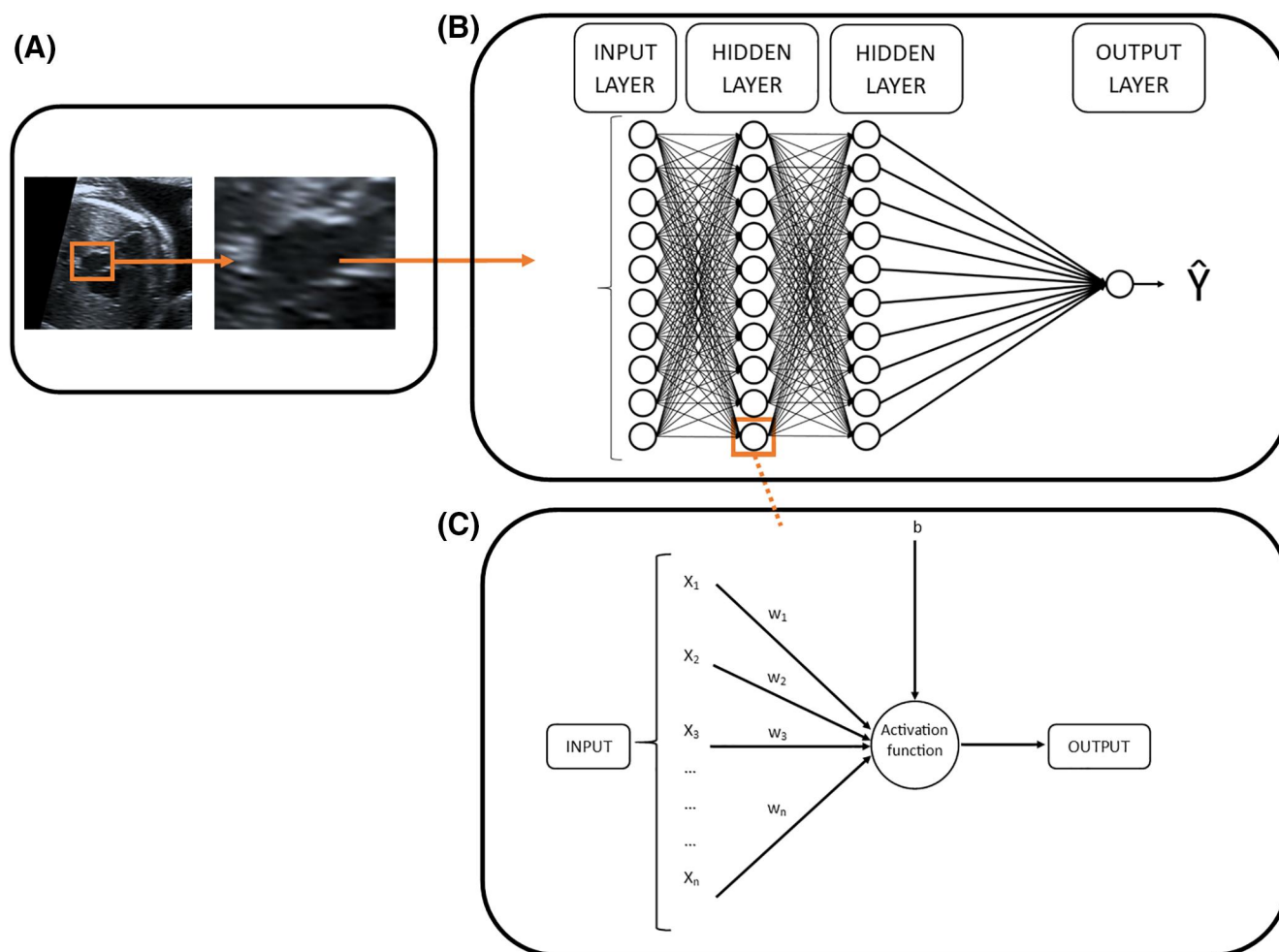


FIGURE 2 How fetal echocardiogram data (A) might be incorporated into a neural network (B), and how the network is composed of multiple perceptrons (C). *b*, bias; *w*, weight; *X*, input data; \hat{y} , prediction [Colour figure can be viewed at wileyonlinelibrary.com]

AI has also shown promise in the context of CHD (the primary focus of fetal cardiology). Using supervised DL, Diller et al.⁴¹ trained a convolutional neural network to discriminate accurately between patients with discordant ventriculoarterial connections post atrial switch procedure, patients with both discordant atrioventricular and ventriculoarterial connections, and normal controls. Such direct image-based diagnosis of CHD has exciting potential in fetal cardiology, although with additional challenges. Taking this further, once the diagnosis is known, AI may be able to extract additional information from imaging data and use this to inform prognosis, or perform some other more complex categorization task. As an example of this (although using magnetic resonance data rather than ultrasound), AI has been shown to have utility in the automatic determination of long-term prognosis in patients with repaired tetralogy of Fallot.⁴² And the potential for AI in this context is not only limited to imaging data. A recent study has shown that by combining clinical data, text from clinical letters (using text mining algorithms) and data from other clinical investigations such as electrocardiography and exercise testing, ML algorithms can be trained to estimate prognosis in a large cohort of adult CHD patients.⁴³ As

well as predicting outcome, these algorithms can also predict medical interventions, such as the commencement of specific drug therapy. Although not in current clinical use, there is the clear potential in the future for such algorithms in not only predicting, but suggesting, medical treatments and interventions.

4 | INTELLIGENT IMAGING OF THE FETUS

Table 1 outlines some examples of AI methods and how they have been recently utilized in the field of fetal ultrasound imaging. Given the suboptimal detection of fetal cardiac defects at the screening anomaly ultrasound scan, there may be potential for AI to improve this. AI may also be useful to the fetal cardiac specialist, as although the performance of fetal echocardiography is excellent in expert hands, there are still potential routes to AI improving this performance further. However, fetal cardiac ultrasound imaging is a challenging and complex task. There is a high degree of operator dependency, meaning that different operators may not produce images that are similar in appearance, despite representing the same

TABLE 1 AI methods used in some recent key papers using AI in the field of fetal cardiac ultrasound imaging

AI method	Reference	Training data	Application of AI method
Random forests	Bridge et al. ⁵⁷	91 Fetal echocardiogram clips from 12 fetuses	Automatic detection of fetal cardiac position, orientation (axis) and phase
	Le et al. ⁶⁹	Echocardiograms from 3910 fetuses (14.1% with CHD)	Differentiation of fetuses with normal hearts from those with a variety of congenital heart defects
Fully connected neural network	Sulas et al. ⁷⁰	LV inflow–outflow PW Doppler traces from 5 healthy fetuses	Automatic identification of E, A, and V waves on PW Doppler traces
Recurrent neural networks	Chen et al. ⁵¹	Ultrasound scans of 300 fetuses	Automatic detection of standard planes, including the four-chamber cardiac view
Convolutional neural networks	Dong et al. ⁵⁸	2032 Examples of fetal four-chamber views and 5000 views of other fetal structures	Automatic detection of the four-chamber cardiac view, and an automatic assessment of image quality
	Baumgartner et al. ⁵⁹	Ultrasound scans of 1003 healthy fetuses	Automatic detection of standard planes, including the four-chamber, LV outflow, RV outflow, and three-vessel view cardiac views
	Arnaout et al. ⁶⁸	107,832 Images from 1326 ultrasound scans of fetuses (400 with CHD)	Automatic identification of standard cardiac planes, then differentiation between normal hearts and those with a variety of congenital heart defects. Also automatic segmentation of fetal heart structures to allow biometric measurement

Abbreviations: AI, artificial intelligence; CHD, congenital heart disease; LV, left ventricle; PW, pulsed wave; RV, right ventricle.

anatomic area. The operator can vary several parameters such as gain, contrast, resolution, depth, and magnification meaning that the images are not consistent between studies, even of the same patient, and this is compounded by constant fetal movement relative to the ultrasound probe. Imaging artefacts such as acoustic shadowing from bone are common and difficult to avoid, and have the potential to confuse algorithms. Because of the small fetal size, the fetal heart takes up a relatively small proportion of the image (certainly when compared to postnatal echocardiography), meaning that any algorithm will need to learn to ignore a large proportion of the available data. A further difference from postnatal echocardiography is that the orientation and position of the heart in the image is highly variable, creating further complexity in image analysis.

One approach to improve image acquisition is the use of automated 2D reconstructions of 3D data volumes. In this technique the operator obtains a 3D volume of ultrasound data, including the fetal heart, for example using spatiotemporal image correlation. It is possible to then manually postprocess this dataset, producing 2D images that replicate the standard planes of the fetal echocardiogram, allowing the operator to examine these for evidence of CHD. To some extent this technique removes the need for the operator to be skilled in the subtle and complex movements of the probe that are necessary for the acquisition of standard cardiac planes. However, in effect this just swaps one problem for another, as the operator now needs to become skilled in the difficult task of manually manipulating the 3D dataset to display to 2D images of interest. Techniques have been proposed to automate this process.^{44–48} These could be described as a “human support system” rather than ML, as the system is reliant on a human indicating where specific cardiac landmarks are, then uses predefined rules to reconstruct the required 2D planes.

Results have been published demonstrating the potential utility of this technique in fetal cardiac screening.⁴⁹ However, although case reports have shown this technique being of use in detecting isolated cases of CHD, no larger trial demonstrating an improvement in antenatal detection of CHD has yet been published.⁵⁰

The acquisition of standard cardiac image planes is the first step in diagnosing CHD from ultrasound images. Using AI to automatically retrieve these planes from a stream of ultrasound imaging data is a potential route improving detection rates. This may have the potential to reduce the “cognitive load” of the sonographer, allowing them to focus on identification of abnormal anatomy, rather than the pausing and saving of standard planes. The automatically retrieved planes may also be of higher quality than the manually obtained ones, which also may improve diagnostic accuracy. Several groups have investigated this method,^{51–56} including some focusing solely on the fetal heart.^{57,58} Our group in collaboration with others has previously published the SonoNet algorithm.^{59,60} This uses a deep convolutional neural network that was trained using labeled routine mid-trimester ultrasound images from 2694 volunteers. This achieves real-time classification of standard screening planes from a continuous stream of ultrasound video data. Using this we can automatically save the required images, without the sonographer having to freeze and manually save the image. These images can then be automatically used in a standardized reporting template, further streamlining workflow.

In addition to automatic plane detection, work has been published on automated fetal biometry.^{61–63} For similar reasons to above, this could conceivably improve anomaly detection rates by freeing up the sonographer from mundane tasks, who may then be more likely to identify abnormal anatomy. Human-level performance in

measurement accuracy has been demonstrated on retrospectively acquired data. In current practice, we use measurements of cardiac structures indexed to the gestational age of the fetus. If fetal size is automatically determined, then such measurements could be more informatively indexed to the fetal size, and indeed even these cardiac measurements could potentially be automated. In addition to 2D biometric measurements, early work has been published on the automatic segmentation of a 3D volume of ultrasound data.⁶⁴ Although currently this work focuses on identifying the entire fetus rather than specific fetal anatomy, it raises exciting future possibilities of being able to automatically segment the complex 3D cardiac structures, which may be of great use in the identification of fetal cardiac disease.

The quality of the acquired ultrasound images has been shown to relate to the likelihood of clinically significant errors, and this is a particular area of concern in fetal ultrasound scanning as it usually involves lone and siloed working, and it is not standard practice to store the entirety of the ultrasound data stream for later review.⁶⁵ AI has been investigated as a means to automate the quality-control process. Wu et al.⁶⁶ used deep convolutional neural networks to assess the quality of abdominal views, based on the relative size of the region of interest, and key anatomical structures within it. Focusing on the cardiac four chamber view, Dong et al.⁵⁸ both automatically extracted this standard plane from the ultrasound data stream, and then used convolutional neural networks to assess the image quality (based on appropriate gain and zoom, and appearance of key anatomical landmarks). This automatic quantification of image quality may be a useful adjunct to standard ultrasound scanning performed by humans, but also will likely form an integral part of any more automated system.

Relatively little work has been published examining the use of AI to directly classify fetal hearts into normal or abnormal. Arnaout et al.^{67,68} have proposed the use of an ensemble of neural networks to explore this. Using 107,823 ultrasound images from 1326 echocardiograms, they successfully used supervised learning to train a model to identify specific cardiac views. They then used a separate model to differentiate between structurally normal hearts and complex CHD (considered as one group). Good performance was achieved, with an area under the curve (AUC) of 0.99, sensitivity of 95%, and specificity of 96%, which is comparable to expert human clinician performance. Le et al.⁶⁹ have also published similar work in abstract form, with good model performance on detecting CHD based on retrospective data. They trained random forest algorithms (a different form of ML from neural networks) to differentiate between normal hearts and a group of congenital heart defects, using echocardiogram data from 3910 fetuses, achieving an AUC of 0.83, sensitivity of 93%, and specificity of 72%. As an alternative approach, Sulas et al.⁷⁰ have investigated the use of neural networks to automatically interpret pulsed-wave Doppler traces on the left ventricular inflow and outflow. They postulate that this might be an alternative method of quantifying cardiac time intervals, given the difficulties in obtaining fetal electrocardiographic data, although the utility of this technique in identifying fetal disease has not been

tested. How to translate all of these findings into community-level screening using prospectively acquired data, and how well this would work, has not yet been assessed.

Current practitioners of fetal cardiology will know that information given to prospective parents about a particular diagnosis is usually fairly generic, and other than a few well-described risk factors for poor outcome or need for early intervention, it is difficult to provide a personalized description of the likely life course for each individual fetus. As has been described above for adult CHD patients, AI may be a means of extracting previously unrecognized information from ultrasound imaging data, and perhaps combining this with data from other sources such as other clinical parameters or fetal magnetic resonance imaging. This could provide a prediction not only of diagnosis, but more granular predictions such as life expectancy, need for urgent intervention, and even quality of life long-term. Such precision medicine could have great potential in the parental counseling process, and inform decisions such as place or mode of delivery of the fetus.

5 | SHOULD WE BE WORRIED? ETHICS AND CONTROVERSIES

AI has now become a common part of our everyday lives. Take-up thus far has been more limited in the medical world, partly because of an understandable degree of caution on the part of clinicians when introducing a novel technique into a high-risk situation. Obstetrics is a particularly litigious branch of medicine, and so it would not be surprising if caution were even higher in this specialty. And indeed, there are good reasons to be wary. If there is an error in an algorithm then the potential for adverse impact is large, either by “missing” cardiac lesions or by increasing false positive diagnoses in a systematic manner, compared to individual human error. Confirming this risk, there have already been cases of commercially available AI tools performing poorly and potentially causing harm to patients, by making incorrect decisions when applied to a real-world scenario.⁷¹ This risk is increased by the fact that it is often difficult to identify exactly why neural networks have made a certain prediction, the so-called “black box” problem.^{27,71} This is a particular challenge in medicine, where it is likely that clinicians and the public will demand a degree of “explainability” before an AI tool is considered acceptable.²⁸ It is also possible to inadvertently introduce unwanted bias into an algorithm, for example creating a model that works better in people of certain races, which would clearly be unacceptable for a medical application.^{28,72}

The vast majority of medical AI research has been performed using retrospectively acquired data, which has been used both for training and testing the algorithms. Although many authors have published very favorable results outlining excellent model performance, very little work has been done examining how to apply these models to a real-world setting, on prospectively acquired data. And even fewer studies have demonstrated an actual clinical benefit to patients by using these new tools.⁷³ Although a model may show superior performance to an expert clinician in a highly controlled test

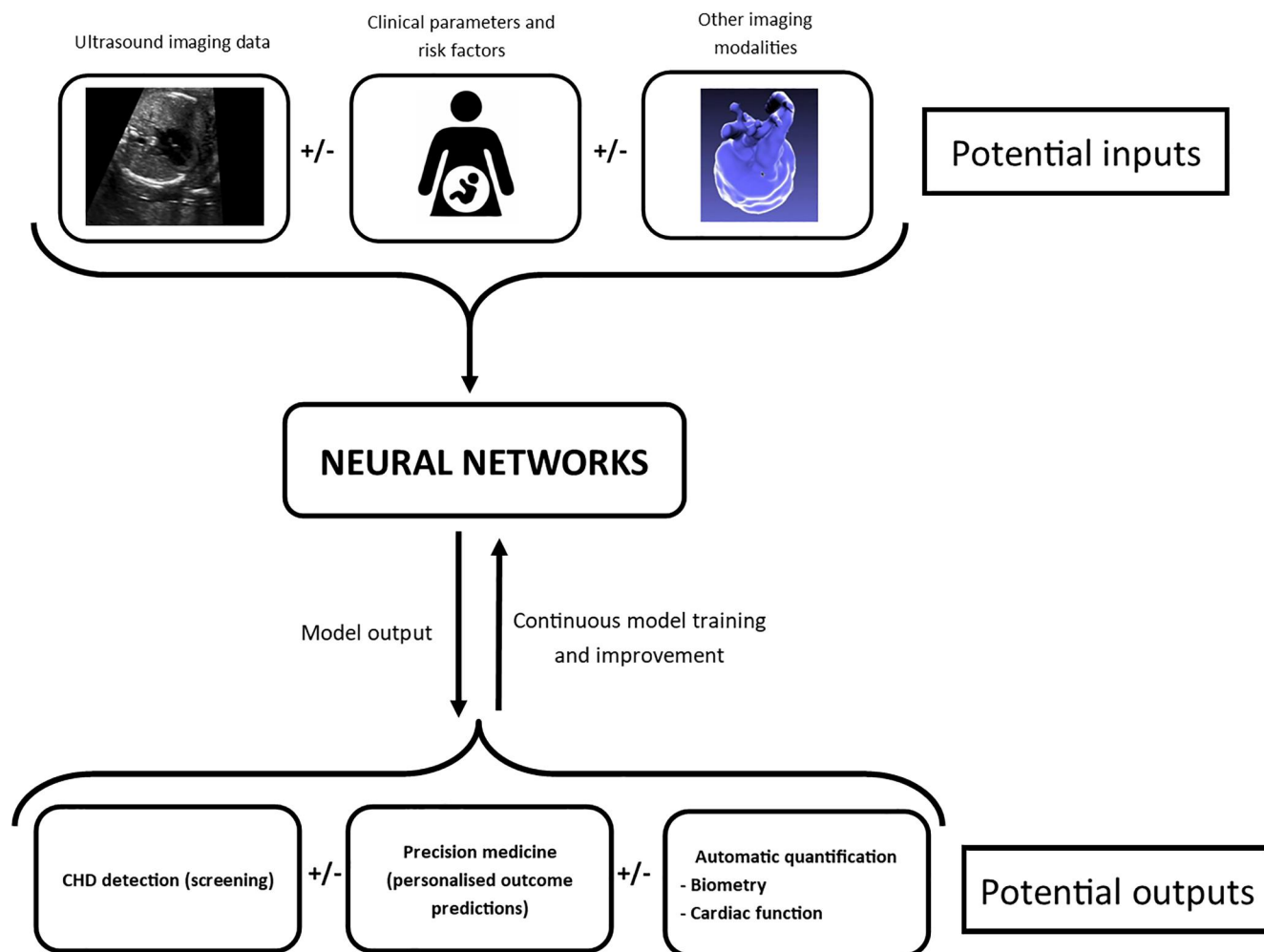


FIGURE 3 Schematic diagram of how machine learning models could be embedded in fetal cardiology clinic workflow [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]

situation using retrospective data, this by no means translates into superior performance in the chaotic and unpredictable world of clinical medicine. It is very difficult to interpret model performance in these terms unless high-quality randomized controlled trials are performed in the future. A recent systematic review found only two published randomized controlled trials comparing the performance of AI versus humans in medical imaging tasks, and identified a high risk of bias in the majority of the published literature using retrospective data.⁷⁴

6 | THE FUTURE

At the present time there is a lack of a robust framework to help design and assess clinical trials of AI, and exactly how these trials should be undertaken is debatable. As we have discussed, there are a variety of ways that AI could be applied to fetal cardiology, ranging from automated measurement of cardiac biometry, through to direct diagnosis of cardiac malformations, with the sonographer using this information to improve their performance. What we are asking of AI

will define how models are developed in the future, and how the clinical trials to assess these models are designed.

If we want to train models to detect individual lesions using a supervised approach, then a large amount of training data will be required. Crucially, for supervised learning this will need to be labeled data. For some nonmedical applications of ML it has been possible to use the general public to provide labeled training data (e.g., clicking on pictures that contain specified items), but clearly this would not be possible in fetal cardiology, where a high degree of expertise is needed, creating a scarcity of labeled data that is an impediment to algorithm training. One scenario would be AI algorithms that are completely embedded in the workflow of the fetal cardiology clinic. In this way, labeled data could be continuously fed to the ML algorithms that are continuously improving in their performance. Work has been done specifically exploring this concept in the context of fetal cardiology, and this may be a means by which AI can both inform and learn simultaneously.⁷⁵ However, the regulatory framework in most countries would not currently allow such an approach to be utilized in clinical practice, as approval so far has only been granted for finalized “locked” algorithms, but not for those that are constantly changing.

"Catastrophic forgetting," where an algorithm abruptly deteriorates in its performance on the original task when a new task is learned, is a major barrier to the implementation of such continuous learning.⁷⁶ Nevertheless, the US Food and Drug Administration last year published a discussion paper outlining how such continuously adapting algorithms might be regulated in future clinical use.^{77,78} Figure 3 outlines how neural networks could be embedded in either screening or specialist fetal cardiology workflows, with the potential inputs and outputs of such models.

There may be other solutions to the problem of labeled data scarcity, such as using semi-supervised approaches (where a model trained on a different task is adapted for a new task by providing a small amount of new training data), or unsupervised methods (where the model sorts the data into similar clusters, then the label is needed only for each cluster). Alternatively, an entirely different approach may be needed where a model is trained to identify "signatures" that are common amongst many types of CHD, and thus identifies fetuses at increased risk. For a truly successful implementation of AI in fetal cardiology, it is likely that novel methods such as these will be required, and this is an ongoing focus of our group's research.

7 | CONCLUSION

AI shows great promise for future application in fetal cardiology. As a specialty, it faces perhaps greater challenges to realize this promise when compared to other branches of medicine, including the scarcity of labeled data due to the rarity and heterogeneity of CHD, and challenges specific to ultrasound imaging. Nevertheless, once these issues are overcome, along with the development of appropriate regulatory and governance frameworks, it is likely that AI will form at least some part of routine fetal cardiac care in the near future. We envisage a future where AI works in tandem with skilled clinicians to optimize performance. Although the replacement of clinicians with algorithms seems a very long way off, it will likely be within our lifetimes that a familiarity with the uses, and pitfalls, of AI will be mandatory for many healthcare professionals.

ACKNOWLEDGMENTS

This work was supported by the Wellcome Trust (IEH Award, 102431), by core funding from the Wellcome/EPSRC 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. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

CONFLICT OF INTERESTS

The authors have stated explicitly that there is no conflict of interests in connection with this article.

ORCID

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

REFERENCES

1. Wang X, Xiao J. The use of fetal echocardiography for pregnancy diagnosis. *Chin J Obstet Gynecol*. 1964;10(4):267-269.
2. Allan LD, Tynan MJ, Campbell S, Wilkinson JL, Anderson RH. Echocardiographic and anatomical correlates in the fetus. *Br Heart J*. 1980;44(4):444-451. <https://doi.org/10.1136/hrt.44.4.444>
3. Kleinman C, Hobbins J, Jaffe C, Lynch D, Talner N. Echocardiographic studies of the human fetus: prenatal diagnosis of congenital heart disease and cardiac dysrhythmias. *Pediatrics*. 1980;65:133-134.
4. Allan LD, Sharland GK, Milburn A, et al. Prospective diagnosis of 1,006 consecutive cases of congenital heart disease in the fetus. *J Am Coll Cardiol*. 1994;23(6):1452-1458. [https://doi.org/10.1016/0735-1097\(94\)90391-3](https://doi.org/10.1016/0735-1097(94)90391-3)
5. Stümpflen I, Stümpflen A, Wimmer M, Bernaschek G. Effect of detailed fetal echocardiography as part of routine prenatal ultrasonographic screening on detection of congenital heart disease. *Lancet*. 1996;348(9031):854-857. [https://doi.org/10.1016/S0140-6736\(96\)04069-X](https://doi.org/10.1016/S0140-6736(96)04069-X)
6. Salomon LJ, Alfrevic Z, Berghella V, et al. Practice guidelines for performance of the routine mid-trimester fetal ultrasound scan. *Ultrasound Obstet Gynecol*. 2011;37(1):116-126. <https://doi.org/10.1002/uog.8831>
7. International Society of Ultrasound in Obstetrics & Gynecology. Cardiac screening examination of the fetus: guidelines for performing the "basic" and "extended basic" cardiac scan. *Ultrasound Obstet Gynecol*. 2006;27(1):107-113. <https://doi.org/10.1002/uog.2677>
8. 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(7):1-12. <https://doi.org/10.1136/bmjopen-2018-028139>
9. Quartermain MD, Pasquali SK, Hill KD, et al. Variation in prenatal diagnosis of congenital heart disease in infants. *Pediatrics*. 2015;136(2):e378-85. <https://doi.org/10.1542/peds.2014-3783>
10. National Congenital Heart Disease Audit. *Summary Report*. London: National Institute for Cardiovascular Outcomes Research; 2019.
11. Mahle WT, Clancy RR, McGaurn SP, Goin JE, Clark BJ. Impact of prenatal diagnosis on survival and early neurologic morbidity in neonates with the hypoplastic left heart syndrome. *Pediatrics*. 2001;107(6):1277-1282. <https://doi.org/10.1542/peds.107.6.1277>
12. Bonnet D, Coltri A, Butera G, et al. Detection of transposition of the great arteries in fetuses reduces neonatal morbidity and mortality. *Circulation*. 1999;99(7):916-918. <https://doi.org/10.1161/01.CIR.99.7.916>
13. Calderon J, Angeard N, Moutier S, et al. Impact of prenatal diagnosis on neurocognitive outcomes in children with transposition of the great arteries. *J Pediatr*. 2012;161(1):94-8.e1. <https://doi.org/10.1016/j.jpeds.2011.12.036>
14. Hunter LE, Simpson JM. Prenatal screening for structural congenital heart disease. *Nat Rev Cardiol*. 2014;11(6):323-334. <https://doi.org/10.1038/nrcardio.2014.34>
15. Holland BJ, Myers JA, Woods CR. Prenatal diagnosis of critical congenital heart disease reduces risk of death from cardiovascular compromise prior to planned neonatal cardiac surgery: a meta-analysis. *Ultrasound Obstet Gynecol*. 2015;45(6):631-638. <https://doi.org/10.1002/uog.14882>
16. Simpson J, Zidere V, Miller O. *Fetal Cardiology—A Practical Approach to Diagnosis and Management*. New York, NY: Springer International Publishing.

17. Nisselrooij AEL, Teunissen AKK, Clur SA, et al. Why are congenital heart defects being missed? *Ultrasound Obstet Gynecol*. 2019;747-757. <https://doi.org/10.1002/uog.20358>
18. Tegnander E, Eik-Nes SH. The examiner's ultrasound experience has a significant impact on the detection rate of congenital heart defects at the second-trimester fetal examination. *Ultrasound Obstet Gynecol*. 2006;28(1):8-14. <https://doi.org/10.1002/uog.2804>
19. Uzun O, Kennedy J, Davies C, et al. Training: improving antenatal detection and outcomes of congenital heart disease. *BMJ Open Qual*. 2018;7(4), e000276. <https://doi.org/10.1136/bmjoq-2017-000276>
20. Sharland GK, Allan LD. Screening for congenital heart disease prenatally. Results of a 2 1/2-year study in the South East Thames Region. *Br J Obstet Gynaecol*. 1992;99(3):220-225. <https://doi.org/10.1111/j.1471-0528.1992.tb14503.x>
21. Hunter S, Heads A, Wyllie J, Robson S. Prenatal diagnosis of congenital heart disease in the northern region of England: benefits of a training programme for obstetric ultrasonographers. *Heart*. 2000;84(3):294-298. <https://doi.org/10.1136/heart.84.3.294>
22. Krittanawong C, Johnson KW, Rosenson RS, et al. Deep learning for cardiovascular medicine: a practical primer. *Eur Heart J*. 2019;40(25):2058-2069C. <https://doi.org/10.1093/eurheartj/ehz056>
23. Kulikowski CA. Beginnings of artificial intelligence in medicine (AIM): computational artifice assisting scientific inquiry and clinical art—with reflections on present AIM challenges. *Yearb Med Inform*. 2019;28(1):249-256. <https://doi.org/10.1055/s-0039-1677895>
24. Mitchell T. *Machine Learning*. New York: McGraw Hill Education; 1997.
25. Bizopoulos P, Koutsouris D. Deep learning in cardiology. *IEEE Rev Biomed Eng*. 2019;12(c):168-193. <https://doi.org/10.1109/RBME.2018.2885714>
26. www.deeplearning.ai
27. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *N Engl J Med*. 2019;380(14):1347-1358. <https://doi.org/10.1056/NEJMra1814259>
28. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med*. 2019;17(1):1-9. <https://doi.org/10.1186/s12916-019-1426-2>
29. Hwang UJ, Park S, Jin K-N, et al. Development and validation of a deep learning-based automatic detection algorithm for active pulmonary tuberculosis on chest radiographs. *Clin Infect Dis*. 2019;69(5):739-747.
30. Nam JG, Park S, Hwang EJ, et al. Development and validation of deep learning-based automatic detection algorithm for malignant pulmonary nodules on chest radiographs. *Radiology*. 2019;290(1):218-228. <https://doi.org/10.1148/radiol.2018180237>
31. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542(7639):115-118. <https://doi.org/10.1038/nature21056>
32. McKinney SM, Sieniek M, Godbole V, et al. International evaluation of an AI system for breast cancer screening. *Nature*. 2020;577(7788):89-94. <https://doi.org/10.1038/s41586-019-1799-6>
33. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*. 2016;316(22):2402-2410. <https://doi.org/10.1001/jama.2016.17216>
34. Lopez-jimenez F, Attia Z, Arruda-olson AM, et al. Artificial intelligence in cardiology: present and future. *Mayo Clin Proc*. 2020;95(May):1015-1039. <https://doi.org/10.1016/j.mayocp.2020.01.038>
35. Alsharqi M, Woodward WJ, Mumith JA, Markham DC, Upton R, Leeson P. Artificial intelligence and echocardiography. *Echo Res Pract*. 2018;5(4):R115-R125. <https://doi.org/10.1530/erp-18-0056>
36. Madani A, Arnaout R, Mofrad M, Arnaout R. Fast and accurate view classification of echocardiograms using deep learning. *npj Digit Med*. 2018;1(1):1-8. <https://doi.org/10.1038/s41746-017-0013-1>
37. Zhang J, Gajjala S, Agrawal P, et al. Fully automated echocardiogram interpretation in clinical practice: feasibility and diagnostic accuracy. *Circulation*. 2018;138(16):1623-1635. <https://doi.org/10.1161/CIRCULATIONAHA.118.034338>
38. Knackstedt C, Bekkers SCAM, Schummers G, et al. Fully automated versus standard tracking of left ventricular ejection fraction and longitudinal strain the FAST-EFs multicenter study. *J Am Coll Cardiol*. 2015;66(13):1456-1466. <https://doi.org/10.1016/j.jacc.2015.07.052>
39. Medvedofsky D, Mor-Avi V, Amzulescu M, et al. Three-dimensional echocardiographic quantification of the left-heart chambers using an automated adaptive analytics algorithm: multicentre validation study. *Eur Heart J Cardiovasc Imaging*. 2018;19(1):47-58. <https://doi.org/10.1093/ehjci/jew328>
40. Tsang W, Salgo IS, Medvedofsky D, et al. Transthoracic 3D echocardiographic left heart chamber quantification using an automated adaptive analytics algorithm. *JACC Cardiovasc Imaging*. 2016;9(7):769-782. <https://doi.org/10.1016/j.jcmg.2015.12.020>
41. Diller GP, Babu-Narayan S, Li W, et al. Utility of machine learning algorithms in assessing patients with a systemic right ventricle. *Eur Heart J Cardiovasc Imaging*. 2019;20(8):925-931. <https://doi.org/10.1093/ehjci/jez211>
42. Diller GP, Orwat S, Vahle J, et al. Prediction of prognosis in patients with tetralogy of Fallot based on deep learning imaging analysis. *Heart*. 2020;106:1-8. <https://doi.org/10.1136/heartjnl-2019-315962>
43. Diller GP, Kempny A, Babu-Narayan SV, et al. Machine learning algorithms estimating prognosis and guiding therapy in adult congenital heart disease: data from a single tertiary centre including 10 019 patients. *Eur Heart J*. 2019;40(13):1069-1077. <https://doi.org/10.1093/eurheartj/ehy915>
44. Abuhamad A. Automated multiplanar imaging: a novel approach to ultrasonography. *J Ultrasound Med*. 2004;23(5):573-576. <https://doi.org/10.7863/jum.2004.23.5.573>
45. Abuhamad A, Falkensammer P, Reichartseder F, Zhao Y. Automated retrieval of standard diagnostic fetal cardiac ultrasound planes in the second trimester of pregnancy: a prospective evaluation of software. *Ultrasound Obstet Gynecol*. 2008;31(1):30-36. <https://doi.org/10.1002/uog.5228>
46. Yeo L, Romero R. Fetal Intelligent Navigation Echocardiography (FINE): a novel method for rapid, simple, and automatic examination of the fetal heart. *Ultrasound Obstet Gynecol*. 2013;42(3):268-284. <https://doi.org/10.1002/uog.12563>
47. Yeo L, Romero R. Intelligent navigation to improve obstetrical sonography. *Ultrasound Obstet Gynecol*. 2016;47(4):403-409. <https://doi.org/10.1002/uog.12562>
48. Chaoui R, Martins J, Heling S. Recent development in three and four dimension fetal echocardiography. *Fetal Diagn Ther*. 2020;47:345-353. <https://doi.org/10.1159/000500454>
49. Garcia M, Yeo L, Romero R, et al. Prospective evaluation of the fetal heart using fetal intelligent navigation echocardiography (FINE). *Ultrasound Obstet Gynecol*. 2016;47(4):450-459. <https://doi.org/10.1002/uog.15676>
50. Yeo L, Luewan S, Markush D, Gill N, Romero R. Prenatal diagnosis of dextrocardia with complex congenital heart disease using fetal intelligent navigation echocardiography (FINE) and a literature review. *Fetal Diagn Ther*. 2018;43(4):304-316. <https://doi.org/10.1159/000468929>
51. Chen H, Dou Q, Ni D, et al. Automatic Fetal Ultrasound Standard Plane Detection Using Knowledge Transferred Recurrent Neural Networks. In: Navab N, Hornegger J, Wells WM, Frangi A, eds. *Medical Image Computing and Computer-Assisted Intervention—MICCAI*

2015. Cham, Switzerland: Springer International Publishing; 2015:507-514.
52. Chen H, Ni D, Qin J, et al. Standard plane localization in fetal ultrasound via domain transferred deep neural networks. *IEEE J Biomed Heal Informatics*. 2015;19(5):1627-1636. <https://doi.org/10.1109/JBHI.2015.2425041>
53. Yu Z, Wu L, Ni D, et al. Fetal facial standard plane recognition via deep convolutional neural networks. *Chin J Biomed Eng*. 2017;36(3):267-275. <https://doi.org/10.3969/j.issn.0258-8021.2017.03.002>
54. Lei B, Tan EL, Chen S, et al. Automatic recognition of fetal facial standard plane in ultrasound image via Fisher vector. *PLoS One*. 2015;10(5):85-88. <https://doi.org/10.1371/journal.pone.0121838>
55. Yaqub M, Kelly B, Papageorgiou AT, Noble JA. Guided Random Forests for Identification of Key Fetal Anatomy and Image Categorization in Ultrasound Scans. In: Navab N, Hornegger J, Wells WM, Frangi AF, eds. *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*. Cham, Switzerland: Springer International Publishing; 2015:687-694.
56. Yaqub M, Kelly B, Papageorgiou AT, Noble JA. A deep learning solution for automatic fetal neurosonographic diagnostic plane verification using clinical standard constraints. *Ultrasound Med Biol*. 2017;43(12):2925-2933. <https://doi.org/10.1016/j.ultrasmedbio.2017.07.013>
57. Bridge CP, Ioannou C, Noble JA. Automated annotation and quantitative description of ultrasound videos of the fetal heart. *Med Image Anal*. 2017;36:147-161. <https://doi.org/10.1016/j.media.2016.11.006>
58. Dong J, Liu S, Liao Y, et al. A generic quality control framework for fetal ultrasound cardiac four-chamber planes. *IEEE J Biomed Heal Informatics*. 2020;24(4):931-942. <https://doi.org/10.1109/JBHI.2019.2948316>
59. 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 Imag*. 2017;36(11):2204-2215. <https://doi.org/10.1109/TMI.2017.2712367>
60. Baumgartner CF, Kamnitsas K, Matthew J, Smith S, Kainz B, Rueckert D. Real-Time Standard Scan Plane Detection and Localisation in Fetal Ultrasound Using Fully Convolutional Neural Networks. In: Ourselin S, Joskowicz L, Sabuncu MR, Unal G, Wells W, eds. *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2016*. Cham, Switzerland: Springer International Publishing; 2016:203-211.
61. Kim HP, Lee SM, Kwon JY, Park Y, Kim KC, Seo JK. Automatic evaluation of fetal head biometry from ultrasound images using machine learning. *Physiol Meas*. 2019;40(6). <https://doi.org/10.1088/1361-6579/ab21ac>
62. Li J, Wang Y, Lei B, et al. Automatic fetal head circumference measurement in ultrasound using random forest and fast ellipse fitting. *IEEE J Biomed Heal Informatics*. 2018;22(1):215-223. <https://doi.org/10.1109/JBHI.2017.2703890>
63. Sinclair M, Baumgartner CF, Matthew J, et al. Human-level performance on automatic head biometrics in fetal ultrasound using fully convolutional neural networks. *Annu Int Conf IEEE Eng Med Biol Soc*. 2018;2018:714-717. <https://doi.org/10.1109/EMBC.2018.8512278>
64. Yang X, Yu L, Li S, et al. Towards automated semantic segmentation in prenatal volumetric ultrasound. *IEEE Trans Med Imag*. 2019;38(1):180-193. <https://doi.org/10.1109/TMI.2018.2858779>
65. Dudley NJ, Chapman E. The importance of quality management in fetal measurement. *Ultrasound Obstet Gynecol*. 2002;19(2):190-196. <https://doi.org/10.1046/j.0960-7692.2001.00549.x>
66. Wu L, Cheng JZ, Li S, Lei B, Wang T, Ni D. FUIQA: fetal ultrasound image quality assessment with deep convolutional networks. *IEEE Trans Cybern*. 2017;47(5):1336-1349. <https://doi.org/10.1109/TCYB.2017.2671898>
67. Arnaout R, Curran L, Chinn E, Zhao Y, Moon-Grady A. *Deep-Learning Models Improve on Community-Level Diagnosis for Common Congenital Heart Disease Lesions*. arXiv. <https://arxiv.org/abs/1809.06993>
68. Arnaout R, Curran L, Zhao Y, Levine J, Chinn E, Moon-Grady A. Expert-level prenatal detection of complex congenital heart disease from screening ultrasound using deep learning. *medRxiv*. 2020. <https://doi.org/10.1101/2020.06.22.20137786>
69. Le TK, Truong V, Nguyen-Vo T-H, et al. Application of machine learning in screening of congenital heart diseases using fetal echocardiography. *J Am Coll Cardiol*. 2020;75(11):648. [https://doi.org/10.1016/s0735-1097\(20\)31275-4](https://doi.org/10.1016/s0735-1097(20)31275-4)
70. Sulas E, Ortu E, Raffo L, Urru M, Tumbarello R, Pani D. Automatic recognition of complete atrioventricular activity in fetal pulsed-wave Doppler signals. *Annu Int Conf IEEE Eng Med Biol Soc*. 2018;2018:917-920. <https://doi.org/10.1109/EMBC.2018.8512329>
71. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. 2019;25(1):44-56. <https://doi.org/10.1038/s41591-018-0300-7>
72. Barocas S, Selbst AD. Big data's disparate impact. *SSRN Electron J*. 2018;671:671-732. <https://doi.org/10.2139/ssrn.2477899>
73. Shah NH, Milstein A, Bagley PhD SC. Making machine learning models clinically useful. *JAMA*. 2019;322(14):1351-1352. <https://doi.org/10.1001/jama.2019.10306>
74. 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. <https://doi.org/10.1136/bmj.m689>
75. Patra A, Noble JA. Hierarchical class incremental learning of anatomical structures in fetal echocardiography videos. *IEEE J Biomed Heal Informatics*. 2020;24(4):1046-1058. <https://doi.org/10.1109/JBHI.2020.2973372>
76. Kirkpatrick J, Pascanu R, Rabinowitz N, et al. Overcoming catastrophic forgetting in neural networks. *Proc Natl Acad Sci U.S.A.* 2017;114(13):3521-3526. <https://doi.org/10.1073/pnas.1611835114>
77. Minssen T, Gerke S, Aboy M, Price N, Cohen G. Regulatory responses to medical machine learning. *J Law Biosci*. 2020;1-18. <https://doi.org/10.1093/jlb/lsaa002>
78. FDA. *Proposed regulatory framework for modifications to artificial intelligence/machine learning (AI/ML) -based software as a medical device (SaMD)—discussion paper and request for feedback*. US Food and Drug Administration. 2019:1-20.

How to cite this article: Day TG, Kainz B, Hajnal J, Razavi R, Simpson JM. Artificial intelligence, fetal echocardiography, and congenital heart disease. *Prenatal Diagnosis*. 2021;41:733-742. <https://doi.org/10.1002/pd.5892>