Computers in Biology and Medicine 186 (2025) 109666

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

A comprehensive scoping review on machine learning-based fetal echocardiography analysis

Netzahualcoyotl Hernandez-Cruz [a],*, Olga Patey [b], Clare Teng [a], Aris T. Papageorghiou [b], J. Alison Noble [a]

[a] Institute of Biomedical Engineering, University of Oxford, Old Road Campus Research Building, Oxford, OX3 7DQ, UK
[b] Nuffield Department of Women's & Reproductive Health, University of Oxford, Women's Centre, John Radcliffe Hospital, Oxford, OX3 9DU, UK

ARTICLE INFO

Keywords:
Technical scoping review
Congenital heart defects
Fetal echocardiography
Image analysis
Video analysis
Machine learning

ABSTRACT

Fetal echocardiography (ultrasound of the fetal heart) plays a vital role in identifying heart defects, allowing clinicians to establish prenatal and postnatal management plans. Machine learning-based methods are emerging to support the automation of fetal echocardiographic analysis; this review presents the findings from a literature review in this area. Searches were queried at leading indexing platforms ACM, IEEE Xplore, PubMed, Scopus, and Web of Science, including papers published until July 2023. In total, 343 papers were found, where 48 papers were selected to compose the detailed review. The reviewed literature presents research on neural network-based methods to identify fetal heart anatomy in classification and segmentation modelling. The reviewed literature uses five categorical technical analysis terms: attention and saliency, coarse to fine, dilated convolution, generative adversarial networks, and spatio-temporal. This review offers a technical overview for those already working in the field and an introduction to those new to the topic.

1. Introduction

Congenital heart defects (CHDs) are the leading cause of neonatal mortality due to birth defects [1], and affect up to 1.2% of live births worldwide [2]. Early detection of CHDs during fetal life is important because it allows an appropriate plan to be made, such as birth in a proper centre equipped to deal with the condition. Evidence suggests that infants diagnosed after birth are less likely to survive (both before and after heart surgery) and are more likely to have adverse long-term neurological outcomes than those diagnosed prenatally [3].

Assessment of the fetal heart is an integral part of the routine obstetric scan offered to all pregnant women and usually takes place around 20 weeks as part of the routine anatomical survey [4] using fetal echocardiography. This well-established clinical technique uses prenatal ultrasound imaging to evaluate the fetal heart and nearby connected blood vessels [5,6].

Fetal echocardiography is non-invasive and allows real-time data acquisition; it is reliable, safe, and inexpensive [7]. It has evolved with the introduction of spectral, colour Doppler [8] and three-dimensional (3D) imaging [9]. However, despite all these advantages and forming a core component of routine fetal ultrasound, it has been estimated that prenatal CHD detection is missed in approximately half of all cases [10]. Fetal echocardiography in clinical settings remains highly human-dependent, requiring the operator to know the fetal anatomy, excellent hand-eye coordination to acquire informative images [11], and the ability to recognise heart defects (see Fig. 1). Research suggests that the most frequent reasons a CHD is not diagnosed during a second-trimester routine scan are the incorrect acquisition of echocardiographic images or inadequate interpretation of the images acquired [10].

To overcome these deficiencies in image acquisition, detection, and recognition of CHDs, researchers have been investigating the integration of machine learning-based algorithms into fetal ultrasound examination, showing early promising results in clinical image and video analysis, for instance [12–17] (see also Appendix C for details of the reviewed literature).

Several reviews and surveys have been published, for example, [11,14,16,17] focused on the subject of machine learning-based applications in the medical domain, addressing the topic from a clinical perspective, with a brief discussion of machine learning methodology. Reviews such as [13,18] summarise literature on the automation of echocardiogram analysis, providing technical insights but are primarily centred on the analysis of adult echocardiograms. Others, such as [19], offer an informative review; however, due to its broad scope, it dedicates less than 900 words to discuss technical insight on machine learning techniques for the analysis of fetal heart images and videos, which is the main focus of our review. Reviews like [15] briefly discuss

---

* Corresponding author.
E-mail address: netzahualcoyotl.hernandez-cruz@eng.ox.ac.uk (N. Hernandez-Cruz).

https://doi.org/10.1016/j.compbiomed.2025.109666
Received 21 May 2024; Received in revised form 6 January 2025; Accepted 7 January 2025

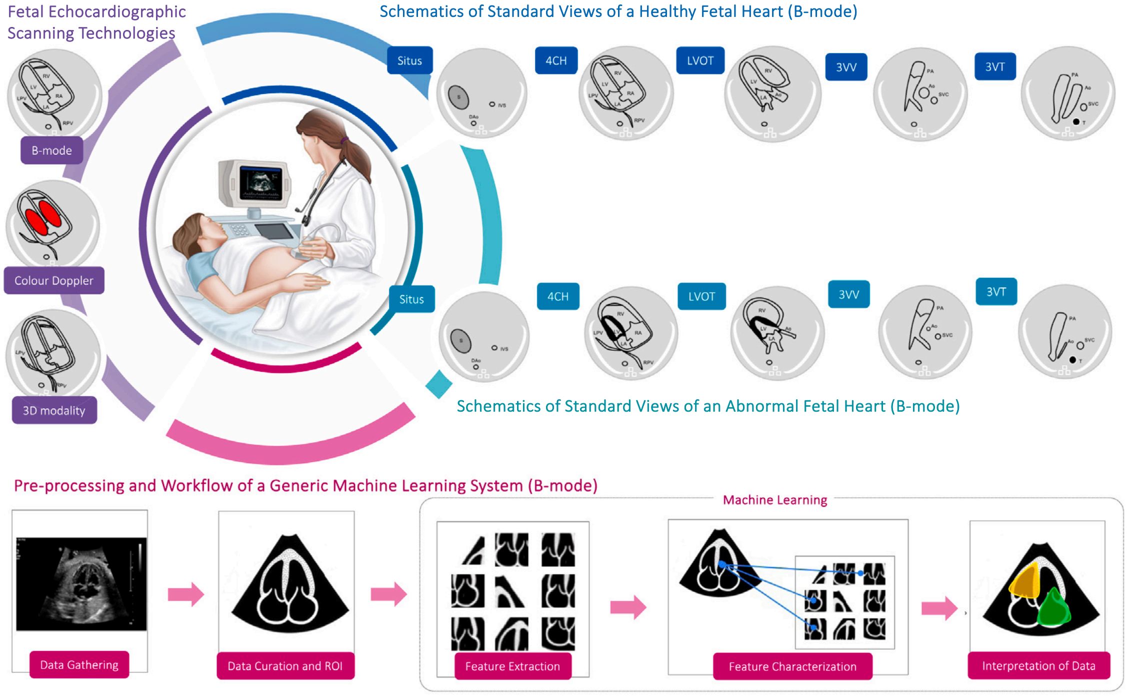

Fig. 1. Overview of implementing machine learning to support clinical decision-making. Echocardiographic acquisition uses three techniques: B-mode, colour Doppler and 3D modalities (Top left). Screening is undertaken by imaging five standard echocardiographic views: situs, four-chamber (4CH), left ventricular outflow tract (LVOT), three-vessel (3VV), three-vessel and trachea (3VT) views (top right). At the bottom, a generic image processing workflow is shown: Data Gathering, as the process of acquiring data, and Data Curation, as the process of enhancing informative image features. Through machine learning-based image analysis (Feature Extraction, Feature Characterisation) and interpretation (Interpretation of Data), automated machine learning-based analysis has the potential to assist in identifying anatomical defects in the fetal heart.

the technical aspects of the machine learning methods. Contrary to the above reviews and surveys, our review paper summaries studies focused on automated fetal echocardiography image and video analysis, emphasising the technical contributions of the reviewed literature. Our review aims to answer the research question, "What neural network machine learning methods and fetal heart anatomical structures have been researched to analyse fetal echocardiographic images and videos?". It adds to the existing literature by providing:

- A technical scoping review of machine learning methods for automated fetal echocardiographic images and video analysis.
- A taxonomy of methods used in studies and the datasets found in the reviewed literature on automated analysis of fetal echocardiographic images and videos.

2. Methodology

The reviewed literature was analysed as a scoping report [20] following Okali and Schabram's guide, a methodology which aims to produce a comprehensive review and discussion of the existing literature in the field of information systems [21].

2.1. Protocol & training

Authors NH, OP, CT, ATP, and JAN collaborated on this scoping review paper. With expertise in computational clinical image analysis, NH, CT, and JAN led the technical contributions. OP, with expertise in fetal cardiology, and AP, with expertise in obstetrics and fetal medicine, led the clinical contributions. Additionally, AN and ATP provided guidance and oversight throughout the research process.

NH conducted the initial literature search across multiple databases, systematically retrieved relevant papers based on pre-determined search terms, and screened titles and abstracts to identify potentially eligible papers. Throughout the review process, NH, OP, and CT assessed the quality and relevance of the included papers using established appraisal criteria. Any disagreements were resolved through discussions.

2.2. Searching the literature

Searches were restricted to journals and conference proceedings with a technological and scientific orientation at ACM, IEEE Xplore, PubMed, Scopus, and Web of Science databases.

The literature search consisted of four groups of keywords: (i) stage of gestation (fetal, fetus, feticide, foetus, embryo, heart, cardiac), (ii) ultrasound modality (echocardiography, echo, ultrasound, doppler), (iii) data type (image, video, 3d, 4d), and (iv) computational analysis approach (machine learning, artificial intelligence, neural network, deep learning, expert system, knowledge engineering, intelligent retrieval) (queries are available at Appendix A). Searches were restricted to papers written in English. There was no restriction on the publication year. The last literature search update was conducted in July 2023.

2.3. Practical screening and quality appraisal

The retrieved papers were cross-referenced to avoid duplication, then screened by title and abstract to ensure they fit within the scope of interest, followed by full-text screening for quality appraisal. Findings from the searched literature identified the early year of 2000 as the beginning of research in fetal echocardiographic image analysis using machine learning, with an increase in published literature since 2010, with the highest number of publications between 2019 – 2023. This coincides with the introduction of CNNs in 1989 [22], the first paper using CNNs for fetal echocardiographic image analysis [23] in 1999, and the AlexNet architecture [24] — first deep CNN referenced a landmark in computer vision in 2012 leading to the adoption of deep convolutional neural networks in computer vision tasks. We also noticed that the keywords expert system, knowledge engineering, and

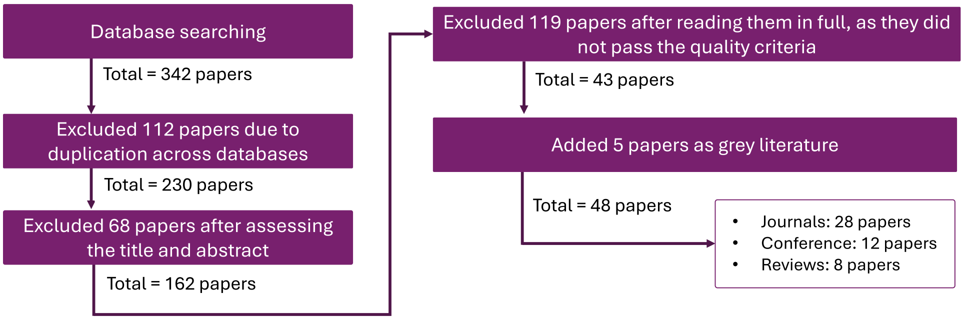

Fig. 2. Flow diagram illustrating the steps in the selection criteria for the papers included in this review.

intelligent retrieval did not contribute to adding literature to our search. Despite their minimal impact, we retained these terms in the search strategy to ensure the query remained broad enough to capture potential interdisciplinary studies, as these concepts are closely related to the broader topic in the group of keywords "(iv) computational analysis approach".

As shown in Fig. 2, a total of 342 papers were retrieved. We excluded 112 papers due to duplication across the databases. Following this, we excluded 68 papers because their titles and abstracts did not suggest using echocardiographic analysis and machine learning-based methods. We read the remaining papers in full, from which we excluded 119 papers for not meeting five quality criteria: (1) the research objective of the study is clear; (2) the research focuses on human medical images or videos; (3) the use of machine learning algorithms is clear; (4) the study includes sufficient details in the methodology, experiment, and results; and (5) the study adds technical value to the existing literature or showcases the applicability of machine learning methods in the medical field. We also excluded papers consisting of research in the adult heart and fetal plane/view detection, which would not contribute to answering our research question or use cases with no methodological or technical contribution. Additionally, we added 5 papers identified from reference lists or otherwise not identified in the original search (grey literature). This review concentrates on 48 relevant papers, i.e., 28 journal research papers, 12 research papers in conference proceedings, and 8 review/survey papers. The relevant papers were read in detail; we selected 20 papers to discuss in our main Section 3 due to their insightful contribution to machine learning-based fetal echocardiography analysis.

2.4. Data extraction

We examined the selected papers, focusing on understanding the technical contribution and the data characteristics used. Information regarding the technical methodology was summarised into four categories: (1) technical methodology, consisting of the machine learning algorithm(s) used or extended to develop the claimed contribution. (2) machine learning tasks used to identify and locate specific areas of the heart where the defects are present; tasks include instance segmentation (IS), object detection (OD) and image classification (IC). (3) results, and (4) contribution.

Information regarding the data characteristics was summarised into five categories to facilitate understanding of the research application and assess the extent to which findings generalise: (1) medical condition, i.e. the specific CHD being examined. (2) fetal echocardiographic standard views. (3) fetal anatomical structure being researched, i.e. the region of the heart analysed. (4) data characteristics details such as sample size. (5) gestational age in weeks, as the stage of pregnancy can have varying implications on development and imaging in CHDs. Refer to Appendix D for a table with the taxonomy of datasets in the reviewed literature.

2.5. Analysis of findings

The information extracted in the previous step of the literature protocol was used for analysis and interpretation by contrasting and identifying technical attributes and similarities in the reviewed literature. Findings indicated that research focuses on improving feature extraction methods for machine learning-based segmentation and classification tasks. Seven categories emerged: architectures, attention and saliency, coarse to fine, dilated convolution, generative adversary network, spatio-temporal, and texture feature.

The category of architectures refers to papers evaluating the performance of state-of-the-art backbones using fetal echocardiographic images, including ensemble architectures. The attention and saliency analysis category relates to methods that include drawing attention to visually salient areas, helping to highlight regions that may contain visual clues, and helping the machine model make a decision. Coarse-to-fine refers to methods where a neural network architecture processes an image by progressively analysing and refining the information at different levels of abstraction. Dilated convolution analysis refers to methods that help generate increased-resolution feature maps, enabling more comprehensive context information extraction without introducing additional parameters to the neural networks. Generative adversarial networks refer to learning deep representations without extensive annotated training data; this enables them to mitigate overfitting and capture data distributions, simplifying image generation tasks. Texture feature analysis refers to converting visual features extracted from images into a structured representation similar to textual data. Spatio-temporal analysis refers to the combined analysis of spatial and temporal dimensions in video.

We decided to omit two categories for detailed discussion: (1) architecture — as this category encompasses known architectures applied to fetal echocardiographic images without modification, and (2) texture feature analysis — as such reported methods were considered not as up-to-date as more recent deep learning methods. These studies are not discussed further but are included in Appendix C for reference and completeness.

3. Technical methodology categorisations

This section organises the reviewed literature according to our five methodology categories. Fig. 3 illustrates how machine learning tasks like instance segmentation of anatomical heart structures, image classification (including plane detection of fetal hearts following the FASP protocol[1]), and object detection of anatomical heart structures, are distributed across these five categories. Refer to Appendix B for a list of acronyms and abbreviations of clinical conditions and terminology used in this section.

---

[1] https://www.gov.uk/government/publications/fetal-anomaly-screening-programme-handbook/20-week-screening-scan

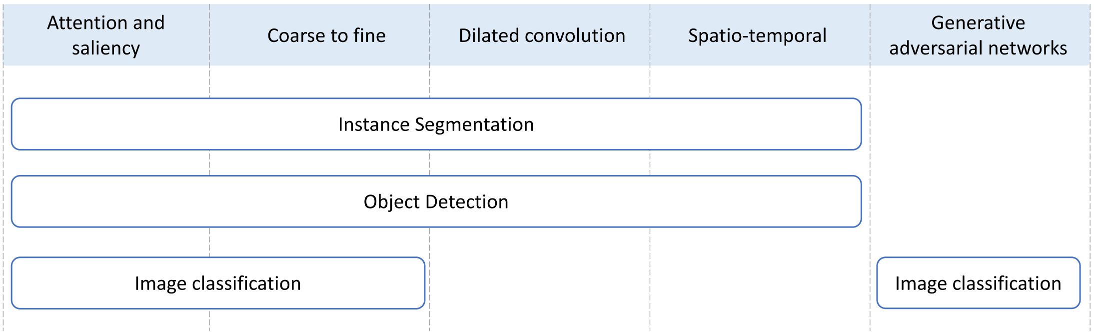

Fig. 3. Summary of instance segmentation, image classification, and object detection tasks across the technical methodology categorisation surveyed in this paper.

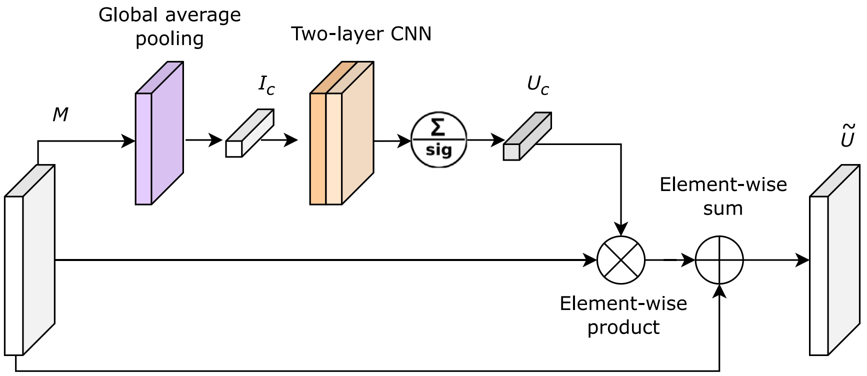

Fig. 4. Illustration of the residual channel attention (RCA) block based on [28].

3.1. *Attention and saliency*

Attention and saliency analysis has been used by researchers in the segmentation and detection of heart structures. These analyses can help identify and visualise important features of an image by employing methods that highlight regions visually salient [25]. In CHD detection, saliency maps can draw attention to regions with abnormalities or findings associated with heart conditions [26]. Researched methods include integrating residual channel attention blocks to explore inter-channel interactions of feature maps [27,28] and using attention maps as a metric for continuously training models [29].

The work in [27] used the segmenting objects by locations (SOLO-v2) [30], an instance segmentation method with category and mask branches. The category branch predicts the instance's class, and the mask branch creates its segmentation mask. The reviewed paper extended SOLO-v2 by adding an attention branch, which takes the fusion of a feature pyramid network (FPN) feature map (levels P2, P3, P4, P5, P6), which contains deep and shallow semantic information as input. The method achieved an average Dice coefficient of 0.78 with an average precision of 0.46 for heart structure segmentation in the 4CH for the clinical conditions NAHS, AVSD, EA, FCR, HLHS, PA/IVS, and TAPVC.

The work in [28] integrated an attention mechanism [25] into their neural network backbone. This enhancement aimed to enhance feature extraction for heart structures while disregarding irrelevant background features. The method includes a residual channel attention (RCA) module and a residual spatial attention (RSA) module to consider spatial context and channel interactions. In the RCA block (Fig. 4), global average pooling forms a channel descriptor I_C. A two-layer CNN and gating mechanism generate a logical channel structure U_C. A weighted sum of U_C and the original feature map M produces a temporary channel attention map. This is added to M via a residual connection to create the final channel attention map $\widetilde{U}$. The RSA block (Fig. 5) uses a convolutional layer to produce a spatial descriptor I_S, which forms a logical spatial structure U_S.

In the RSA block, the spatial information is combined with the input feature maps to create a temporary spatial attention map using a weighted sum. Like the RCA block, the RSA block adds this temporary map to the input feature map. This combined result is passed through a residual connection to produce the final spatial attention feature map $\widetilde{U}$ showing where to highlight or suppress. Results achieved an F1 score of 0.94 for heart structure detection in the 4CH of the NAHS clinical condition.

Other work, like [29], used a class activation mapping with the gradient-weighted class activation mapping (GradCAM) [31] to generate attention maps. The underlined strategy consists of using the attention maps as a metric for assessing the models' ability to learn from unseen classes, thus enabling the adaptation capabilities of models in the lifelong learning [32] approach. For a loss function, the method implements a dual objective of minimising a shift on learnt representations in the form of prior logits in the final model layers using a knowledge distillation approach [33] and performing a cross-entropy classification on the current task classes. The method improved the accuracy of 0.63 for heart structure detection in the 3VT, 3VV, 4CH, and LVOT of the NAHS clinical condition.

3.2. *Coarse to fine*

Coarse-to-fine analysis has been researched for heart structure segmentation. The idea behind coarse-to-fine analysis is to break down a complex analysis into simpler analyses, each focusing on a different level of detail. At the initial coarse level of a network, layers may capture global features; the network refines these features as the data flows through subsequent layers, incorporating more localised and high-level details [34]. Researched methods include combining architectures with different classification strategies [34–36], or joining them with fusion techniques [37].

The work in [35] joined two consecutive U-Nets [38], wherein the initial network incorporates a coarse loss function, and the subsequent

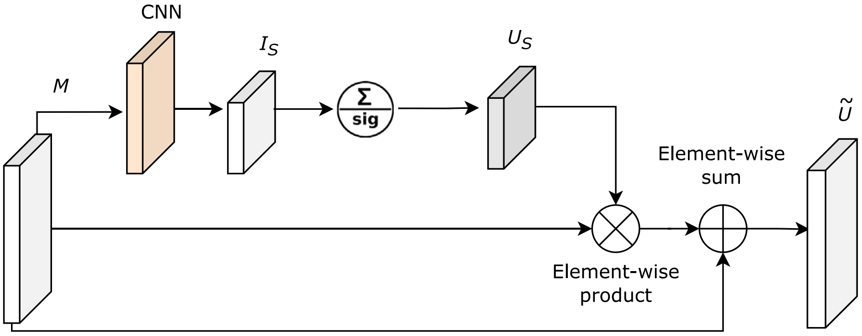

Fig. 5. Illustration of the residual spatial attention (RSA) block based on [28].

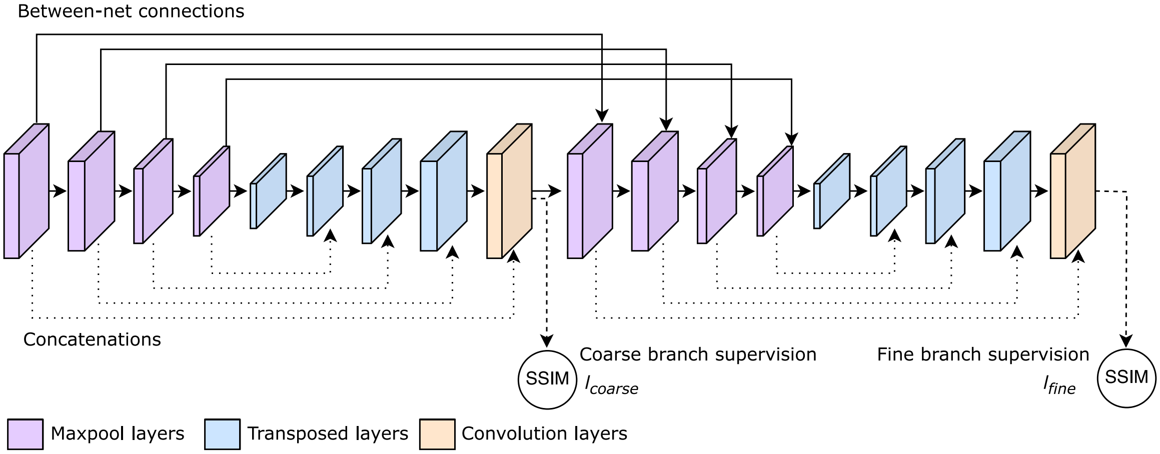

Fig. 6. Illustration of the cascaded of U-Nets based on [35], where each network consists of three types of layers, max-pooling, transposed, and convolutional. The decoder is symmetrical to the encoder, and a skip connection connects the encoder's feature maps with the decoder's feature maps.

network employs a fine loss function (Fig. 6). The loss function consists of the structural similarity index (SSIM) [39], which learns structural information. Each layer from the networks comprises three operations: convolution, instance normalisation, and application of a ReLU activation function. The first network performs a coarse segmentation and sends the extracted features to the second network for further segmentation refinement. To address the inherent problem of vanishing gradients caused by the cascaded networks, the method used a two-fold process: first, by adding auxiliary supervision of the first network, and second, by redesigning an inter-network connection from the encoder layers of the first network to the encoder layers of the second network. The method avoids connecting the networks from the first decoder to prevent prior information of shallow layers from being preserved and transferred to deeper layers, which can affect the characterisation of the heart structure. The method achieved an average Dice coefficient of 0.86, a Hausdorff distance of 3.31, and a pixel accuracy of 0.93 for heart structure segmentation in the A4CH of the NAHS clinical condition.

The work in [36] used a method consisting of two network models (called super and sub) with different classification strategies. The super-model is a coarse classification consisting of convolutional layers leading to a temporal-marching recurrent stage using long short-term memory (LSTM) layers to classify video sequence input into pre-defined top-level classes. The model output comprises SoftMax probabilities, combined and averaged for a batch of inputs to determine the coarse or top-level class. This approach differs from traditional classification pipelines like in [34], where each input is classified separately. The sub-model consists of a fine classification implemented by a set of sub-models representing each coarse class the super-model has been pre-trained with. The two models are pre-trained for classification with convolutional and fully-connected architectures and initialised in a Xavier-improved schema [40]. Results showed an accuracy of 0.72 for heart structure detection in the 3VT, 3VV, 4CH, and LVOT for the clinical conditions NAHS, ASD, ECD, FCR, HLHS, PA/IVS, and TAPVC.

Similarly, the work in [37] designed a method consisting of two interactive branches, coarse and fine, joined over the fusion of features. As shown in Fig. 7, the coarse branch processes the input as a complete image, and the fine branch processes the input in image patches. Branches are designed using dual-path chains as the building block, consisting of a three-layer CNN, followed by a gated axial-transformer encoder inspired by Axial-DeepLab [41], and upsampling layers, where the gated axial-transformer encoders consist of learnable parameters that enable the network to be used for any size data set. The coarse and fine branches simultaneously extract high-level semantic information and spatial dependencies among image patches. Respective features are fused over pixel-wise addition, obtaining precise contour information of the segmented objects. The method uses two dual-path chains to control how much information the positional embedding learns. Results showed an F1 score of 0.97 and an intersection over union (IoU) of 0.94 for heart structures segmentation in the 4CH of the NAHS clinical condition.

3.3. Dilated convolution

Dilated convolution has been researched for heart structure detection and segmentation. It can help generate higher-resolution feature maps, enabling more comprehensive context information extraction without introducing additional parameters to the neural networks. Incorporating dilated convolutions into a network architecture allows it to capture a broader range of spatial dependencies and enhances the model's ability to understand intricate details within the input data [42]. Nevertheless, excessively high dilation rates can result in an

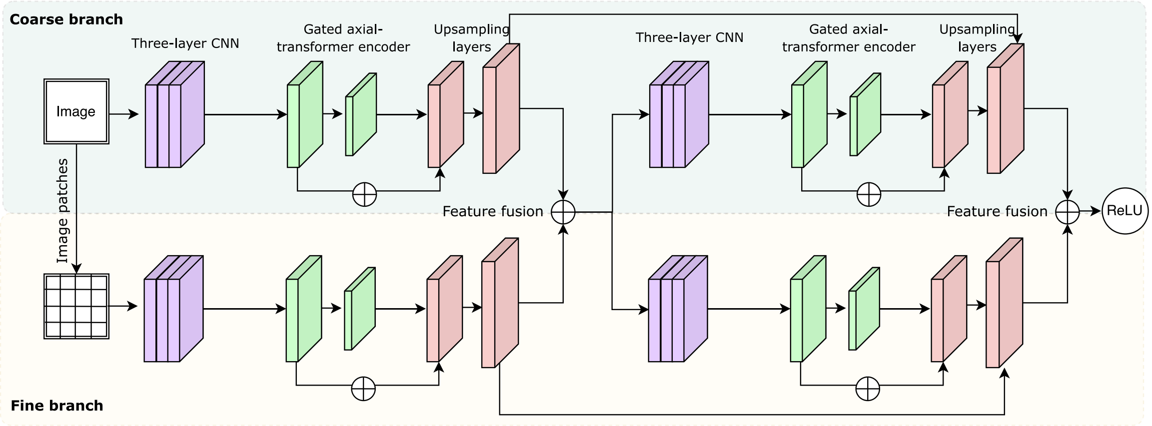

Fig. 7. Illustration of the architecture consisting of four dual-path chains based on [37].

overly receptive field [43], potentially impeding the model to capture local details or subtle variations in the input [44]. Researched methods include strategies like finding the balance between local and global context [44,45], and the use of different dilation rates [42].

Finding an appropriate balance between local and global context is key to ensuring accurate and meaningful representations. For example, the work in [45] used a series of dilated convolutions organised sequentially to achieve the saw-tooth wave-like effect [46] (see Figs. 8 and 9), helpful to expand the receptive field progressively, allowing the network to improve local information utilisation by capturing multi-scale contextual information simultaneously. In addition to the dilated network, the method proposes a fine-grained network consisting of convolutional and transposed convolutional layers inspired by the U-Net architecture. The method mitigates unwanted distortions and enhances the quality of the input data, facilitating the accurate localisation of the targeted heart structures and enabling the segmentation in images with poorly defined boundaries. The method enables the propagation of the underlying information that might be lost in higher-resolution layers, which can delineate heart structure details for more precise region boundary segmentation. The method achieved a Dice coefficient of 0.83, pixel accuracy of 0.93, and AUC of 0.99 for heart structure segmentation in the A4CH of the NAHS clinical condition. These are broadly similar to the results achieved in [44] that used a method to increase the receptive fields and prioritise intricate details using spatial pyramid pooling (SPP) [47] block, which combines multi-scale features by employing multiple parallel filters in a dilated convolution manner. SPP uses dilated convolutions with different dilation rates, which enables the original image with multiple filters to complement the receptive fields, thus capturing useful image context at multiple scales. The latter method suggests using a multi-class dice coefficient as the loss function and then adding binary cross-entropy loss as a constraint on each pixel to improve the performance when implementing SPP. The method achieved an average for IoU of 0.72, cross-entropy (CE) of 0.03, and Dice coefficient of 0.90 for heart structures segmentation in the 4CH for the clinical conditions NAHS, EA, ECD, FCR, HLHS, PA/IVS, and TAPVC.

The work in [42] used dilated convolution with different dilated rates relevant to the heart structures (i.e., larger rates for structures like the heart ventricles and atriums and small rates for structures like the tricuspid and mitral valves) in combination with receptive field blocks (RFB) [48] to enhance the feature representation. The method explores the use of grouped and dilated convolutions in a bottleneck structure with three layers to achieve this. Experiments show that the method improves the feature representation, making the detection of anatomical structures more accurate, reporting a mean average precision (mAP) of 0.94 without compromising inference speed for heart structures detection in the 4CH of the NAHS clinical condition.

3.4. *Generative adversarial networks*

Generative adversarial networks (GANs) have been researched to improve the classification of heart images, such as in-plane detection. GANs are a type of machine learning that simultaneously trains two networks: one for image generation and the other for discrimination. They offer a way to learn deep representations without large amounts of annotated training data, enabling them to capture data distributions [49]. However, GAN training can be challenging, as it requires balancing the competition between the generator and discriminator and mitigating potential issues like mode collapse, where the generator fails to capture the full diversity of the training data, resulting in the generator producing limited or repetitive samples. Researched methods include adopting adversarial cycle learning [50], sub-adversarial losses and attention mechanisms [51].

The work in [50] presented a network referred to as DANomaly, which extends the adversarially learned one-class classifier architecture (ALOCC) [52] by incorporating adversarial cycle learning to train an end-to-end one-class classification network. As shown in Fig. 10, DANomaly consists of the raw and noise image model branches. The raw image branch reconstructs images using the ALOCC architecture to minimise the difference between the raw I and reconstructed images I_O. A sub-adversarial loss is used to enhance the reconstruction ability. The noise image branch reconstructs noise images and emphasises their proximity to the raw images. The Encode–Decode from the noise image branch trains a mapping relationship between the input I_O and output images N_O. Thus, when representative samples train the model, the reconstructed image of the outlier samples will be distorted because the mapping relationship has been changed. When inputting a noise image, the ideal reconstructed image should be close to the raw image, not the noise image itself. DANomaly leverages the raw image model to enhance the noise model's classification capability by improving the reconstruction performance. The various loss functions encourage the relationship between the reconstructed and raw images, resulting in better performance and accuracy. The network is then used with an intact architecture that follows a Wasserstein GAN gradient penalty (WGAN-GP) [53] for classification tasks achieving a performance of 0.85 accuracy and AUC of 0.88 on the reported test data for classification of 4CH images of the NAHS clinical condition.

Similarly, the work in [51] proposed a one-class classification method to recognise out-of-distribution samples corresponding to fetal hearts diagnosed with CHDs. The method consists of an auto-encoding generative adversarial network (α-GAN) model, where the discriminator networks aim to learn the salient features of the heart area during training and generate an out-of-distribution score. As shown in Fig. 11, the method includes two discriminators focusing on the latent feature space and data, respectively. The latent feature space discriminator

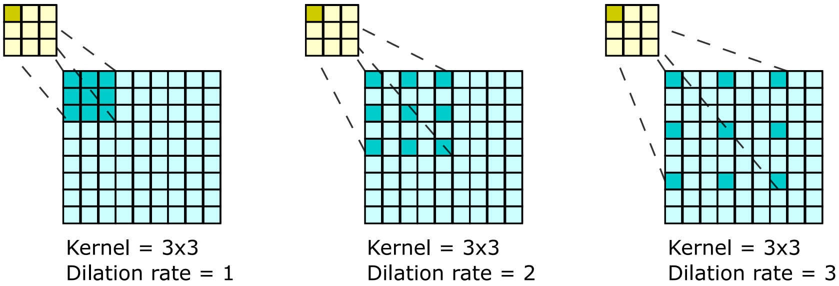

Fig. 8. Illustration exemplifying the dilated convolution at different rates.

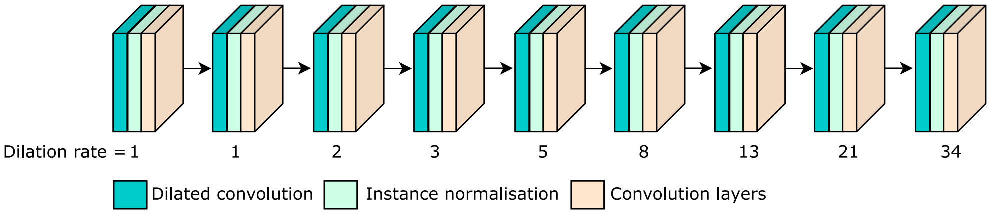

Fig. 9. Illustration exemplifying the series of dilated convolutions organised sequentially, which aims to gain accurate localisation and grasp the global information based on [45].

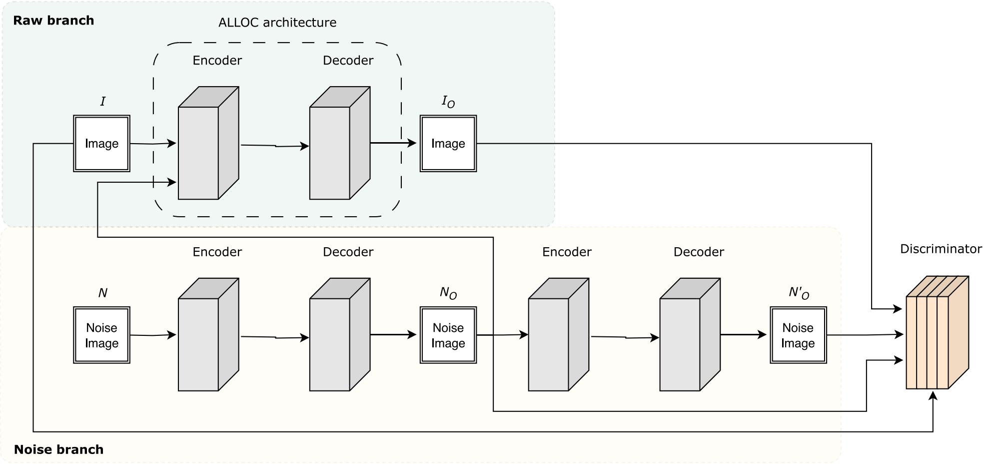

Fig. 10. DANomaly network, which extends the ALOCC architecture based on [50].

consists of four linear layers with Leaky ReLU activation functions. The method discriminates latent representations I_F generated by the encoder network from samples of a standard Gaussian distribution. The data discriminator consists of four convolutional-batch normalisation-ReLU layers. It takes the input image I and the generated image R and outputs a probability of the input being real or fake. The out-of-distribution score is computed using a gradient-based method by applying GradCam to the output of the data discriminator; it also outputs an attention map useful to localise the CHDs. The intuition of using attention maps for computing anomaly scores is based on the hypothesis that after training, the discriminator learns to discriminate between normal and abnormal samples and focuses on relevant features in the image. Thus, in the case of a heart diagnosed with CHD, a discriminator should identify and locate the defects. This work reaches an AUC of 0.81 to classify 4CH images of the HLHS and NAHS clinical conditions.

3.5. *Spatio-temporal*

Spatio-temporal analysis has been researched for heart structure detection and segmentation and the classification of images of the heart, such as in-plane detection. Spatio-temporal refers to the combined analysis of both spatial and temporal dimensions in the context of video (moving images and motion patterns); it poses challenges due to the complexity of accurately interpreting video and tracking motion patterns [54]. Various methods have been reported, including adapting recurrent neural networks (RNNs), extending YOLO-based networks, and fusing data [55–57,57–59].

For example, the work in [55] used RNNs within an architecture design to propagate temporal information across video frames. The method replaces the first layer from each sliding window of a VGG-16 architecture with a recurrent layer. The recurrent layer is a bi-directional RNN consisting of two separate LSTM hidden layers that process the input sequence forward and backward. The inputs to each LSTM layer contain convolutional features computed from the current window at the current frame and a temporal vector from the previous frame. The bi-directional RNN allows for two goals: first, it allows each prediction to be conditioned on the same spatial location over sequential frames; second, it works at a local, regional level of the frames and can focus on local details of the heart, making it discriminative for describing heart details. They achieved a location error of 0.22 and a class error of 0.16 for the heart structures detection in 3VV, 4CH, and LVOT of the NAHS clinical condition. Other works like [56–58] used RNNs to detect the fetal heart and identify the cardiac cycles as a classification task. The method in [57] integrated

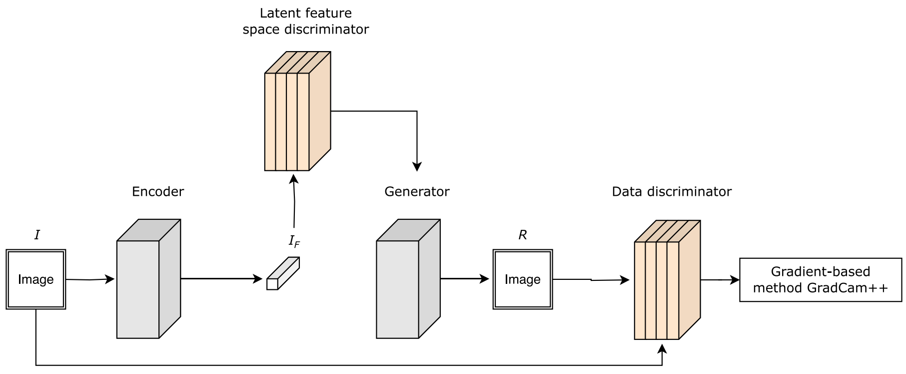

Fig. 11. Architecture of the α-GAN model based on [51].

three modules, first to extract the fetal heart based on target detection using the YOLO-v3 [60] architecture, second to retain temporal dependency information, and third to add a classification strategy for localising the phases of the cardiac cycle. The first module uses the YOLO-v3 architecture. The second consists of fusion methods used to retrain temporal dependency between images. The classification module consists of three-category classes, including the cardiac cycle phases (end-systolic and end-diastolic) and a middle phase representing those images between phases, achieving an accuracy up to 0.95 frame mismatch for the end-systolic and end-diastolic frames on correct samples of classification of heart structures in A4CH, B4CH, P4CH of the NAHS clinical condition.

Similar to the application of neural networks for cardiac cycle detection, [56] proposed a neural network consisting of two concatenated networks, where training data is in the form of linear interpolation of the heart region to a sinusoidal curve as $\sin(y_i)$, where $y_i = \frac{\pi}{2}$ and $y_i = \frac{3\pi}{2}$ represented the end-systolic and end-diastolic phases of the cardiac cycle. The method used a feature extractor network of six sequential convolutional layers alternated with a pooling operation and batch normalisation layers. Logits from the first network concatenate to a second consisting of four sequential RNNs, where each RNN, in turn, returns the hidden state of the network and passes it onto the next layer. The final layer then takes a channel-wise average of the hidden states to determine the cardiac cycle phase, reporting a mean squared error concerning the human-annotated ground truth of 0.18 for the classification of 4CH images of the HLHS and NAHS clinical conditions. Contrary to sinusoidal interpolation, in [58], this method used the variance between consecutive images to quantify motion. Specifically, the method uses an optic flow network [61] trained using gradients of image sequences to train a model to detect cardiac motion. The method used the architecture of FlowNet [62], which enables two branches, one consisting of simple convolutional layers that process an image pair to extract the motion information between them and a second one that performs multiplicative patch comparisons feature maps to differentiate between the end-systolic and end-diastolic phases of the cardiac cycle. The paper presents qualitative results to identify the heart structure motion in the PLAX of the NAHS clinical condition.

In addition to fetal heart detection, [59] focused on segmenting heart structures via a YOLO-v3-based architecture. The proposed method differs from the classical two-stage (object detection and mask prediction) instance segmentation (YOLO-v3) in that the segmentation results of the two-stage approach are either overly dependent on the detector's performance or severe inter-class competition in mask prediction. The model extends the traditional YOLO-v3 architecture with a one-stage instance segmentation branch where the prediction of the mask does not depend on the object detection task; the design includes a new multi-level non-maximum suppression (NMS) mechanism to improve further the segmentation performance that consists of three levels of selection. The output is a binary mask. As the specific tasks of foreground determination, classification, localisation, and segmentation focus on different content, the method decouples the four tasks as four branches, ensuring that each task branch achieves more accurate predictions. The intuition behind the model's strategy is that the object detection step's accuracy heavily influences the quality of the segmentation results. If the detector fails to detect an object accurately or misses it altogether, the subsequent mask prediction will also be compromised. This decoupling allows the model to learn more specialised features for each task, avoiding potential confusion or errors that might arise when these tasks are combined. Results report an accuracy of 0.84 for heart structure segmentation in the 4CH of the NAHS clinical condition.

4. Discussion and future directions

In this section, we discuss and present insight into future directions on five interconnected topics: data accessibility, video analysis, data dimensionality, first trimester of gestation, and Doppler echocardiography.

4.1. Data accessibility

Machine learning-based analysis requires large and diverse datasets for training and testing. However, various factors hinder large-scale data and annotation access, such as privacy laws and data regulations,[2] acquisition or access costs, data heterogeneity due to multi-vendor systems and standardisation of acquisition, and the rareness of certain health conditions [2,63]. For instance, in the reviewed literature, none of the selected papers made their data sets available for open-access. In addition to this, we are only aware of two open access data sets[3],[4] of fetal heart echocardiography images and videos [64,65]. The open-access data is available in portable network graphics (PNG) and joint photographic group (JPG) format.

Throughout the reviewed literature, we noted common practices included using JPG and PNG for image format, and MPEG-4 (MP4) and audio video interleave (AVI) formats for videos. These formats are common due to ultrasound machines' anonymisation and export

---

[2] https://www.gov.uk/data-protection

[3] https://zenodo.org/records/3904280

[4] https://sites.google.com/view/ped-life/data-sets

options. While digital imaging and communications in medicine (DICOM) files are the standard formats for medical imaging, current ultrasound machines used in clinical practice do not offer DICOM files anonymisation.

Another challenge with ultrasound data is that it is derived from different acquisition settings (medical vendor equipment and data management software). Neglecting these variations when designing machine learning models can lead to performance issues and reduced generalisability of the models.

Future directions to address the lack of data in fetal echocardiography analysis could include synthetic data, self-supervised learning [66, 67], and federated analysis [68]. Data augmentation aims to generate synthesised samples to augment a dataset. This augmentation technique has been widely utilised and demonstrated to enhance model performance by increasing the diversity and variability of the training data, mitigating heterogeneity and imbalance properties of some datasets, and thus improving generalisation [66]. Algorithm-level techniques include self-supervised methods that leverage unlabelled data to learn useful representations or features without requiring explicit supervision or labelled examples [67]. Federated analysis may help develop fetal echocardiography image and video analysis due to the opportunities to establish collaborations across different institutions without violating privacy regulations and is particularly relevant due to the rarity of some CHDs [68].

Federate learning has been investigated in applications such as brain imaging [69–71], pulmonary imaging-related diseases [69,72, 73], among others (see references within [74]); nevertheless, there seems to be scarcity evidence on the specific implementation of fetal heart ultrasound analysis, and particularly in the topic of CHD [68].

4.2. Video availability

In clinical practice, storing fetal echocardiographic video may be done for different purposes, such as reviewing to confirm the diagnosis, reviewing discrepancies in diagnosis after birth/on postmortem, auditing, and teaching. However, due to storage costs, video is not stored in routine scanning in many clinical settings. The primary clinical focus is often on immediate diagnosis, usually prioritising the documentation of key findings, measurements, and diagnostic reports that can be achieved through image storage rather than comprehensive storage of video.

Research indicates potential benefits of automated echocardiographic video analysis, such as simplifying scanning acquisition protocols [75], allowing the visualisation and assessment of the fetal heart structure and function, blood flow, and breathing movements [76]. The dynamic evaluation of fetal heart function can then support the monitoring of the processes of maturation and development of the fetal central nervous system, which has been proposed as a method for differentiating between normal and abnormal growth [77].

Future research could focus on leveraging the potential benefits of fetal echocardiographic video analysis, as the reviewed literature indicates that research has not yet successfully developed machine-learning-based video analysis methods to address CHD detection in real-world scenarios effectively. While automated video-based analysis shows promise, further advances are needed to harness its full potential in fetal medicine. The use of efficient data storage formats and libraries for efficiently managing large N-dimensional arrays[5] may support and enable the effective handling of large datasets, similar to the use of medical imaging frameworks such as SimpleITK[6] and MONAI.[7]

4.3. Data dimensionality

Fetal echocardiograms can represent fetal heart anatomy in 2D (width and height), 3D (width, height, and depth), and 4D (3D with the addition of time, capturing motion or changes over time).

While 2D imaging in fetal analysis has been widely adopted as the standard visualisation for heart structures for diagnosis, it has limitations in providing in-depth information and accurately representing the inter-planar relationships between anatomical structures for specific medical conditions, potentially leading to misinterpretation, as it happens on other medical domains [78]. Matrix transducers [79] offer the ability to capture 3D information directly [80]. However, the skills and techniques that lead to successful decision-making in the 2D echocardiograms are not the same as those that will lead to success in the reading of 3D echocardiograms; this is true not only for the sonographer [81] but also for machine learning algorithms [82], likely due to the reduced information available in 2D compared with 3D data [83].

Research using 3D and 4D data in fetal echocardiography represents a promising future direction due to the potential benefits of acquiring a more comprehensive representation of fetal heart anatomy. Current literature is limited; our literature search retrieved only three papers with preliminary insight into machine learning-based methods applied to 3D [84] and 4D [58,85] data. Papers [58,85] focused on optical flow algorithms for object recognition and motion detection; their research demonstrated that the obtained optical flow information could represent the movement pattern of the heart. The paper [84] fused extracting features (colour and texture) from 3D images to train deep neural networks for image classification tasks.

4.4. First-trimester of gestation

The detection of cardiac abnormalities represents a distinct challenge for prenatal screening. In many countries, the gold standard involves a second-trimester evaluation of cardiac anatomy because it is not before week 10 that the fetal heart is fully developed [86]. This coincides with the findings reported in this review consisting of 31% and 38% of the papers discussing solutions in the second and early third trimester of gestation, respectively (31% of the papers did not disclose the gestational age); see Appendix D for further details.

We argue that future research should investigate the potential of machine learning-based methods in the first trimester, as clinical studies in the first trimester have shown relevant results in fetal echocardiography studies within the range of 11-14 weeks of pregnancy [87]; other citations can be found within [86]. Also, researchers have investigated the inferences between the development of the heart and the placenta, finding them interlinked through several mechanisms, representing an opportunity to predict abnormalities, such for an example as transpositions of the great arteries, which may arise if the placenta fails to function as a selective barrier to xenobiotics and other teratogens in the umbilical-placental circulation, influencing cardiac development in the fetus [88]. This is yet to be demonstrated in the scope of machine learning image and video analysis, which, if successful, could contribute towards detecting fetal heart abnormalities at the first weeks of gestation when the heart is not fully developed.

4.5. Doppler echocardiography

In the clinic, colour Doppler is essential in fetal echocardiography because it identifies blood flow patterns within the heart and blood vessels, which is crucial for detecting CHDs [4]. Colour Doppler images and videos can improve diagnostic accuracy in identifying abnormal blood flow and heart function. Research in this regard, however, is limited. Our literature search found only two papers [58,89] reporting preliminary results when using colour Doppler images. The latter paper used colour Doppler images for view classification and the former for

---

[5] https://zarr.dev/

[6] https://simpleitk.org/

[7] https://monai.io/index.html

motion characterisation of the heart.

Research in colour Doppler techniques is a promising future direction to enhance diagnostic accuracy, minimise inter-observer variability, and improve the prediction of clinical outcomes in fetal echocardiography. This focus could lead to more reliable and efficient assessments of fetal heart conditions.

5. Final remarks

This review presents the current state-of-the-art in machine learning-based fetal echocardiography analysis. Findings suggested active research on machine learning-based methods in five categories: attention and saliency, coarse to fine, dilated convolution, generative adversarial networks, and spatio-temporal analysis.

We conclude that machine learning is emerging with promising early results in automated fetal echocardiographic analysis. However, further research is needed to demonstrate that machine learning-based methods are effective at scale in real-world scenarios and clinical settings; this includes infrastructure requirements, cost implications, training needs for medical staff, and integration with existing healthcare systems. Future research may also delve deeper into suggestions on policy or framework changes that might be needed to facilitate the adoption of these technologies in healthcare systems.

CRediT authorship contribution statement

Netzahualcoyotl Hernandez-Cruz: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Olga Patey:** Writing – original draft, Validation, Methodology, Data curation, Conceptualization. **Clare Teng:** Writing – original draft, Formal analysis, Data curation. **Aris T. Papageorghiou:** Writing – review & editing, Validation, Supervision, Funding acquisition, Conceptualization. **J. Alison Noble:** Writing – review & editing, Writing – original draft, Supervision, Investigation, Funding acquisition, Conceptualization.

Ethics statement

The requirements for informed written consent and ethical approval were waived because of the nature of the study (i.e., scoping review).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was partly supported by the InnoHK-funded Hong Kong Centre for Cerebro-cardiovascular Health Engineering (COCHE) Project 2.1 (Cardiovascular risks in early life and fetal echocardiography). Co-authors AN and ATP were supported by the Oxford Partnership Comprehensive Biomedical Research Centre with funding from the NIHR Biomedical Research Centre (BRC), UK funding scheme.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.compbiomed.2025.109666.

References

[1] P.T. Levy, A.R. Thomas, A. Wethall, D. Perez, M. Steurer, M.K. Ball, Rethinking congenital heart disease in preterm neonates, NeoReviews (ISSN: 1526-9906) 23 (6) (2022) e373–e387, http://dx.doi.org/10.1542/NEO.23-6-E373.

[2] W. Wu, J. He, X. Shao, Incidence and mortality trend of congenital heart disease at the global, regional, and national level, 1990–2017, Medicine (ISSN: 15365964) 99 (23) (2020) http://dx.doi.org/10.1097/MD.0000000000020593.

[3] J. Burns, A. Basken, R. Acosta, M. Garnier-Villarreal, A. Kulkarni, D.A. Hayes, Identification of research priorities in CHD: empowering patients and families through participation in the development of formal research agendas, in: Cardiology in the Young, Cambridge University Press, (ISSN: 1047-9511) 2022, pp. 1–6, http://dx.doi.org/10.1017/S104795112200066X.

[4] J.S. Carvalho, R. Axt-Fliedner, R. Chaoui, J.A. Copel, B.F. Cuneo, D. Goff, L.G. Kopylov, K. Hecher, W. Lee, A.J. Moon-Grady, H.A. Mousa, H. Munoz, D. Paladini, F. Prefumo, E. Quarello, J. Rychik, B. Tutschek, M. Wiechec, S. Yagel, ISUOG practice guidelines (updated): fetal cardiac screening, Ultrasound. Obstet. Gynecol (ISSN: 14690705) 61 (2023) 788–803, http://dx.doi.org/10.1002/uog.26224.

[5] T.G. Day, B. Kainz, J. Hajnal, R. Razavi, J.M. Simpson, Artificial intelligence, fetal echocardiography, and congenital heart disease, Prenat. Diagn. (ISSN: 1097-0223) 41 (6) (2021) 733–742, http://dx.doi.org/10.1002/PD.5892.

[6] M.T. Donofrio, A.J. Moon-Grady, L.K. Hornberger, J.A. Copel, M.S. Sklansky, A. Abuhamad, B.F. Cuneo, J.C. Huhta, R.A. Jonas, A. Krishnan, S. Lacey, W. Lee, E.C. Michelfelder, G.R. Rempel, N.H. Silverman, T.L. Spray, J.F. Strasburger, W. Tworetzky, J. Rychik, Diagnosis and treatment of fetal cardiac disease: a scientific statement from the American Heart Association, Circulation (ISSN: 1524-4539) 129 (2014) 2183–2242, http://dx.doi.org/10.1161/01.CIR.0000437597.44550.5D.

[7] J. Zhou, M. Du, S. Chang, Z. Chen, Artificial intelligence in echocardiography: detection, functional evaluation, and disease diagnosis, Cardiovasc. Ultrasound (ISSN: 14767120) 19 (1) (2021) http://dx.doi.org/10.1186/S12947-021-00261-2.

[8] M. Meola, J. Ibeas, G. Lasalle, I. Petrucci, Basics for performing a high-quality color Doppler sonography of the vascular access, J. Vasc. Access (ISSN: 17246032) 22 (1) (2021) 18–31, http://dx.doi.org/10.1177/11297298211018060.

[9] B. Behboodi, H. Rivaz, S. Lalondrelle, E. Harris, Automatic 3D ultrasound segmentation of uterus using deep learning, in: IEEE International Ultrasonics Symposium, IUS, IEEE Computer Society, (ISSN: 19485727) 2021, http://dx.doi.org/10.1109/IUS52206.2021.9593671.

[10] A.E. van Nisselrooij, A.K. Teunissen, S.A. Clur, L. Rozendaal, E. Pajkrt, I.H. Linskens, L. Rammeloo, J.M. van Lith, N.A. Blom, M.C. Haak, Why are congenital heart defects being missed? Ultrasound Obstet. Gynecol. (ISSN: 1469-0705) 55 (6) (2020) 747–757, http://dx.doi.org/10.1002/UOG.20358.

[11] M. Gembicki, D.R. Hartge, C. Dracopoulos, J. Weichert, Semiautomatic fetal intelligent navigation echocardiography has the potential to aid cardiac evaluations even in less experienced hands, J. Ultrasound Med. Off. J. Am. Inst. Ultrasound Med. (ISSN: 1550-9613) 39 (2) (2020) 301–309, http://dx.doi.org/10.1002/JUM.15105.

[12] A. Gudigar, U. Raghavendra, J. Samanth, A. Vasudeva, A.J. Ashwal, K. Nayak, R.S. Tan, E.J. Ciaccio, C.P. Ooi, P.D. Barua, F. Molinari, U.R. Acharya, Role of four-chamber heart ultrasound images in automatic assessment of fetal heart: A systematic understanding, Informatics (ISSN: 22279709) 9 (2) (2022) http://dx.doi.org/10.3390/informatics9020034.

[13] Z. Ye, Y.J. Kumar, G.O. Sing, J. Zhang, X. Ni, Deep echocardiography: a first step toward automatic cardiac disease diagnosis using machine learning, J. Internet Technol. (ISSN: 20794029) 21 (6) (2020) 1589–1600, http://dx.doi.org/10.3966/160792642020112106002.

[14] K.Y. Leung, Applications of advanced ultrasound technology in obstetrics, Diagnostics (ISSN: 20754418) 11 (7) (2021) http://dx.doi.org/10.3390/DIAGNOSTICS11071217/S1.

[15] P. Garcia-Canadilla, S. Sanchez-Martinez, F. Crispi, B. Bijnens, Machine learning in fetal cardiology: What to expect, Fetal Diagn. Ther. (ISSN: 1015-3837) 47 (5) (2020) 363–372, http://dx.doi.org/10.1159/000505021.

[16] C.D. Reddy, S. Van den Eynde, S. Kutty, Artificial intelligence in perinatal diagnosis and management of congenital heart disease, Semin. Perinatol. (ISSN: 1558075X) 46 (4) (2022) http://dx.doi.org/10.1016/j.semperi.2022.151588.

[17] F. He, Y. Wang, Y. Xiu, Y. Zhang, L. Chen, Artificial Intelligence in Prenatal Ultrasound Diagnosis, in: Frontiers in Medicine, vol. 8, Frontiers Media S.A., (ISSN: 2296858X) 2021, http://dx.doi.org/10.3389/fmed.2021.729978.

[18] G. Zamzmi, L.Y. Hsu, W. Li, V. Sachdev, S. Antani, Harnessing machine intelligence in automatic echocardiogram analysis: Current status, limitations, and future directions, IEEE Rev. Biomed. Eng. (ISSN: 1941-1189) 14 (2021) 181–203, http://dx.doi.org/10.1109/RBME.2020.2988295.

[19] M.C. Fiorentino, F.P. Villani, M. Di Cosmo, E. Frontoni, S. Moccia, A review on deep-learning algorithms for fetal ultrasound-image analysis, Med. Image Anal. (ISSN: 1361-8415) 83 (2023) 102629, http://dx.doi.org/10.1016/j.media.2022.102629.

[20] Z. Munn, M.D. Peters, C. Stern, C. Tufanaru, A. McArthur, E. Aromataris, Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach, BMC Med. Res. Methodol. (ISSN: 14712288) 18 (1) (2018) 1–7, http://dx.doi.org/10.1186/S12874-018-0611-X/TABLES/1.

[21] C. Okoli, A guide to conducting a standalone systematic literature review, Commun. Assoc. Inf. Syst. (ISSN: 1529-3181) 37 (2015) 43, http://dx.doi.org/10.17705/1CAIS.03743.

[22] Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proc. IEEE 86 (11) (1998) 2278–2324, http://dx.doi.org/10.1109/5.726791.

[23] I.C. on Image Processing, E...t.. its Applications University of Manchester, Fetal echocardiographic image segmentation using neural networks, in: Seventh International Conference on Image Processing and Its Applications, Vol. 2, 2, Institution of Electrical Engineers, London, (ISSN: 0537-9989) ISBN: 0852967179, 1999.

[24] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, Commun. ACM (ISSN: 0001-0782) 60 (6) (2017) 84—-0, http://dx.doi.org/10.1145/3065386.

[25] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, in: I. Guyon, U.V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (Eds.), Advances in Neural Information Processing Systems, Vol. 30, Curran Associates, Inc., 2017.

[26] A. Alqaraawi, M. Schuessler, P. Weiß, E. Costanza, N. Berthouze, Evaluating saliency map explanations for convolutional neural networks: a user study, in: Proceedings of the 25th International Conference on Intelligent User Interfaces, IUI '20, Association for Computing Machinery, New York, NY, USA, ISBN: 9781450371186, 2020, pp. 275–285, http://dx.doi.org/10.1145/3377325.3377519.

[27] S. An, H. Zhu, Y. Wang, F. Zhou, X. Zhou, X. Yang, Y. Zhang, X. Liu, Z. Jiao, Y. He, A category attention instance segmentation network for four cardiac chambers segmentation in fetal echocardiography, Comput. Med. Imaging Graph. (ISSN: 18790771) 93 (2021) http://dx.doi.org/10.1016/j.compmedimag.2021.101983.

[28] S. Qiao, S. Pang, G. Luo, S. Pan, T. Chen, Z. Lv, FLDS: An intelligent feature learning detection system for visualizing medical images supporting fetal four-chamber views, IEEE J. Biomed. Health Inf. (ISSN: 21682208) 26 (2022) 4814–4825, http://dx.doi.org/10.1109/JBHI.2021.3091579.

[29] A. Patra, J.A. Noble, Incremental learning of fetal heart anatomies using interpretable saliency maps, in: Y. Zheng, B.M. Williams, K. Chen (Eds.), Medical Image Understanding and Analysis, Springer International Publishing, Cham, ISBN: 978-3-030-39343-4, 2020, pp. 129–141.

[30] X. Wang, R. Zhang, T. Kong, L. Li, C. Shen, SOLOv2: dynamic and fast instance segmentation, in: Proceedings of the 34th International Conference on Neural Information Processing Systems, NIPS '20, Curran Associates Inc., Red Hook, NY, USA, ISBN: 9781713829546, 2020.

[31] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-CAM: Visual explanations from deep networks via gradient-based localization, in: Proceedings of the IEEE International Conference on Computer Vision, Vol. 2017, Institute of Electrical and Electronics Engineers Inc., (ISSN: 15505499) ISBN: 9781538610329, 2017, pp. 618–626, http://dx.doi.org/10.1109/ICCV.2017.74.

[32] C.X. Ling, T. Bohn, A deep learning framework for lifelong machine learning, 2021, arXiv:2105.00157.

[33] Z. Li, P. Xu, X. Chang, L. Yang, Y. Zhang, L. Yao, X. Chen, When object detection meets knowledge distillation: A survey, IEEE Trans. Pattern Anal. Mach. Intell. (2023) 1–25, http://dx.doi.org/10.1109/TPAMI.2023.3257546.

[34] T. Yang, Y. Zhang, M. Zhu, Y. Wang, S. An, X. Gu, X. Liu, J. Han, Y. He, H. Zhu, Segmentation of ten fetal heart components with coarse-to-fine cascading and dynamic feature powering, IET Image Process. (ISSN: 17519667) 16 (2022) 3831–3841, http://dx.doi.org/10.1049/ipr2.12597.

[35] L. Xu, M. Liu, J. Zhang, Y. He, Convolutional-neural-network-based approach for segmentation of apical four-chamber view from fetal echocardiography, IEEE Access (ISSN: 21693536) 8 (2020) 80437–80446, http://dx.doi.org/10.1109/ACCESS.2020.2984630.

[36] A. Patra, J.A. Noble, Hierarchical class incremental learning of anatomical structures in fetal echocardiography videos, IEEE J. Biomed. Health Inf. (ISSN: 21682208) 24 (4) (2020) 1046–1058, http://dx.doi.org/10.1109/JBHI.2020.2973372.

[37] S. Qiao, S. Pang, G. Luo, Y. Sun, W. Yin, S. Pan, Z. Lv, DPC-MSGATNet: dual-path chain multi-scale gated axial-transformer network for four-chamber view segmentation in fetal echocardiography, in: Complex and Intelligent Systems, Springer International Publishing, (ISSN: 21986053) 2023, http://dx.doi.org/10.1007/s40747-023-00968-x.

[38] O. Ronneberger, P. Fischer, T. Brox, N. Navab, J. Hornegger, W.M. Wells, A.F. Frangi, U-net: Convolutional networks for biomedical image segmentation, in: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III, 1st ed. 2015, in: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, vol. 9351, Springer International Publishing, Cham, (ISSN: 0302-9743) ISBN: 3-319-24573-2, 2015.

[39] H. Zhao, O. Gallo, I. Frosio, J. Kautz, Loss functions for image restoration with neural networks, IEEE Trans. Comput. Imaging 3 (1) (2017) 47–57, http://dx.doi.org/10.1109/TCI.2016.2644865.

[40] H.A. Omar, J.S. Domingos, A. Patra, R. Upton, P. Leeson, J.A. Noble, Quantification of cardiac bull's-eye map based on principal strain analysis for myocardial wall motion assessment in stress echocardiography, in: 2018 IEEE 15th International Symposium on Biomedical Imaging, ISBI 2018, 2018, pp. 1195–1198, http://dx.doi.org/10.1109/ISBI.2018.8363785.

[41] H. Wang, Y. Zhu, B. Green, H. Adam, A. Yuille, L.-C. Chen, Axial-DeepLab: Stand-alone axial-attention for panoptic segmentation, in: A. Vedaldi, H. Bischof, T. Brox, J.-M. Frahm (Eds.), Computer Vision, ECCV 2020, Springer International Publishing, Cham, ISBN: 978-3-030-58548-8, 2020, pp. 108–126.

[42] J. Dong, S. Liu, Y. Liao, H. Wen, B. Lei, S. Li, T. Wang, A generic quality control framework for fetal ultrasound cardiac four-chamber planes, IEEE J. Biomed. Health Inf. (ISSN: 21682208) 24 (4) (2020) 931–942, http://dx.doi.org/10.1109/JBHI.2019.2948316.

[43] W. Luo, Y. Li, R. Urtasun, R. Zemel, Understanding the effective receptive field in deep convolutional neural networks, in: D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, R. Garnett (Eds.), Advances in Neural Information Processing Systems, Vol. 29, Curran Associates, Inc., 2016.

[44] T. Yang, J. Han, H. Zhu, T. Li, X. Liu, X. Gu, X. Liu, S. An, Y. Zhang, Y. Zhang, Y. He, Segmentation of five components in four chamber view of fetal echocardiography, in: 2020 IEEE 17th International Symposium on Biomedical Imaging, ISBI, 2020, pp. 1962–1965, http://dx.doi.org/10.1109/ISBI45749.2020.9098726.

[45] L. Xu, M. Liu, Z. Shen, H. Wang, X. Liu, X. Wang, S. Wang, T. Li, S. Yu, M. Hou, J. Guo, J. Zhang, Y. He, DW-Net: A cascaded convolutional neural network for apical four-chamber view segmentation in fetal echocardiography, Comput. Med. Imaging Graph. (ISSN: 18790771) 80 (2020) http://dx.doi.org/10.1016/j.compmedimag.2019.101690.

[46] C. Li, Z. Qiu, X. Cao, Z. Chen, H. Gao, Z. Hua, Hybrid dilated convolution with multi-scale residual fusion network for hyperspectral image classification, Micromachines (ISSN: 2072-666X) 12 (5) (2021) http://dx.doi.org/10.3390/mi12050545.

[47] X. Tian, X. Liu, T. Zhang, J. Ye, W. Zhang, L. Zhang, X. Shi, F. Fu, Z. Li, C. Xu, Effective electrical impedance tomography based on enhanced encoder–decoder using atrous spatial pyramid pooling module, IEEE J. Biomed. Health Inf. (2023) 1–11, http://dx.doi.org/10.1109/JBHI.2023.3265385.

[48] S. Liu, D. Huang, a. Wang, Receptive field block net for accurate and fast object detection, in: Proceedings of the European Conference on Computer Vision, ECCV, 2018.

[49] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial networks, Commun. ACM (ISSN: 0001-0782) 63 (11) (2020) 139—-144, http://dx.doi.org/10.1145/3422622.

[50] Y. Gong, Y. Zhang, H. Zhu, J. Lv, Q. Cheng, H. Zhang, Y. He, S. Wang, Fetal congenital heart disease echocardiogram screening based on dgacnn: Adversarial one-class classification combined with video transfer learning, IEEE Trans. Med. Imaging (ISSN: 1558254X) 39 (4) (2020) 1206–1222, http://dx.doi.org/10.1109/TMI.2019.2946059.

[51] E. Chotzoglou, T. Day, J. Tan, J. Matthew, D. Lloyd, R. Razavi, J. Simpson, B. Kainz, Learning normal appearance for fetal anomaly screening: Application to the unsupervised detection of hypoplastic left heart syndrome, Mach. Learn. Biomed. Imaging (ISSN: 2766-905X) 1 (2021) 1–25, http://dx.doi.org/10.59275/j.melba.2021-g4dg.

[52] M. Sabokrou, M. Khalooei, M. Fathy, E. Adeli, Adversarially learned one-class classifier for novelty detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2018.

[53] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, A.C. Courville, Improved training of wasserstein GANs, in: I. Guyon, U.V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (Eds.), Advances in Neural Information Processing Systems, Vol. 30, Curran Associates, Inc., 2017.

[54] C.K. Wikle, A. Zammit-Mangion, Statistical deep learning for spatial and spatiotemporal data, Annu. Rev. Stat. Appl. (ISSN: 2326-8298) 10 (1) (2023-03-10).

[55] W. Huang, C.P. Bridge, J.A. Noble, A. Zisserman, Temporal HeartNet: Towards human-level automatic analysis of fetal cardiac screening video, in: M. Descoteaux, L. Maier-Hein, A. Franz, P. Jannin, D.L. Collins, S. Duchesne (Eds.), Medical Image Computing and Computer-Assisted Intervention, MICCAI 2017, Springer International Publishing, Cham, ISBN: 978-3-319-66185-8, 2017, pp. 341–349.

[56] L.H. Lee, J.A. Noble, Automatic determination of the fetal cardiac cycle in ultrasound using spatio-temporal neural networks, in: Proceedings - International Symposium on Biomedical Imaging, Vol. 2020, IEEE Computer Society, (ISSN: 19458452) ISBN: 9781538693308, 2020, pp. 1937–1940, http://dx.doi.org/10.1109/ISBI45749.2020.9098613.

[57] B. Pu, N. Zhu, K. Li, S. Li, Fetal cardiac cycle detection in multi-resource echocardiograms using hybrid classification framework, Future Gener. Comput. Syst. (ISSN: 0167739X) 115 (2021) 825–836, http://dx.doi.org/10.1016/j.future.2020.09.014.

[58] Z. Wang, G. Li, J. Zhou, P. O.Ogunbona, Optical flow networks for heartbeat estimation in 4D ultrasound images, in: ACM International Conference Proceeding Series, Association for Computing Machinery, ISBN: 9781450389501, 2021, pp. 127–131, http://dx.doi.org/10.1145/3467707.3467725.

[59] Y. Lu, K. Li, B. Pu, Y. Tan, N. Zhu, A YOLOX-based deep instance segmentation neural network for cardiac anatomical structures in fetal ultrasound images, IEEE/ACM Trans. Comput. Biol. Bioinform. (ISSN: 15579964) (2022) http://dx.doi.org/10.1109/TCBB.2022.3222356.

[60] J. Redmon, A. Farhadi, YOLOv3: An incremental improvement, 2018, arXiv:1804.02767.

[61] T.-W. Hui, X. Tang, C.C. Loy, A lightweight optical flow CNN —Revisiting data fidelity and regularization, IEEE Trans. Pattern Anal. Mach. Intell. (ISSN: 0162-8828) 43 (8) (2021-8-1).

[62] A. Dosovitskiy, P. Fischer, E. Ilg, P. Hausser, C. Hazirbas, V. Golkov, P. van der Smagt, D. Cremers, T. Brox, FlowNet: Learning optical flow with convolutional networks, in: Proceedings of the IEEE International Conference on Computer Vision, ICCV, 2015.

[63] L. Wang, S. Xu, X. Wang, Q. Zhu, Addressing class imbalance in federated learning, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 35, No. 11, Association for the Advancement of Artificial Intelligence, ISBN: 9781713835974, 2021, pp. 10165–10173, http://dx.doi.org/10.1609/aaai.v35i11.17219.

[64] X.P. Burgos-Artizzu, D. Coronado-Gutiérrez, B. Valenzuela-Alcaraz, E. Bonet-Carne, E. Eixarch, F. Crispi, E. Gratacós, Evaluation of deep convolutional neural networks for automatic classification of common maternal fetal ultrasound planes, Sci. Rep. (ISSN: 2045-2322) 10 (1) (2020) 1–12, http://dx.doi.org/10.1038/s41598-020-67076-5.

[65] C. Stoean, N. Bacanin, W. Paja, R. Stoean, D. Iliescu, C. Patru, R. Nagy, Semantic segmentation of fetal heart components in second trimester echocardiography, Procedia Comput. Sci. (ISSN: 1877-0509) 207 (2022) 3085–3092, http://dx.doi.org/10.1016/j.procs.2022.09.366, Knowledge-Based and Intelligent Information & Engineering Systems: Proceedings of the 26th International Conference KES2022.

[66] S. Qiao, S. Pan, G. Luo, S. Pang, T. Chen, A.K. Singh, Z. Lv, A pseudo-siamese feature fusion generative adversarial network for synthesizing high-quality fetal four-chamber views, IEEE J. Biomed. Health Inf. (ISSN: 21682208) (2022) http://dx.doi.org/10.1109/JBHI.2022.3143319.

[67] C. Yu, S. Li, D. Ghista, Z. Gao, H. Zhang, J.D. Ser, L. Xu, Multi-level multi-type self-generated knowledge fusion for cardiac ultrasound segmentation, Inf. Fusion (ISSN: 15662535) 92 (2023) 1–12, http://dx.doi.org/10.1016/j.inffus.2022.11.004.

[68] N. Hernandez-Cruz, P. Saha, M.M.K. Sarker, J.A. Noble, Review of federated learning and machine learning-based methods for medical image analysis, Big Data Cogn. Comput. (ISSN: 2504-2289) 8 (9) (2024) http://dx.doi.org/10.3390/bdcc8090099, URL https://www.mdpi.com/2504-2289/8/9/99.

[69] D.C. Nguyen, Q.-V. Pham, P.N. Pathirana, M. Ding, A. Seneviratne, Z. Lin, O. Dobre, W.-J. Hwang, Federated learning for smart healthcare: A survey, ACM Comput. Surv. (ISSN: 0360-0300) 55 (3) (2022) 1–37, http://dx.doi.org/10.1145/3501296.

[70] W. Li, F. Milletarì, D. Xu, N. Rieke, J. Hancox, W. Zhu, M. Baust, Y. Cheng, S. Ourselin, M.J. Cardoso, A. Feng, Privacy-preserving federated brain tumour segmentation, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 11861 LNCS, Springer, (ISSN: 16113349) ISBN: 9783030326913, 2019, pp. 133–141, http://dx.doi.org/10.1007/978-3-030-32692-0{_}16/FIGURES/3.

[71] U.C. Srivastava, A. Singh, K.S. Kumar, Intracranial Hemorrhage Detection Using Neural Network Based Methods With Federated Learning, 2020, http://dx.doi.org/10.48550/arxiv.2005.08644.

[72] D.C. Nguyen, M. Ding, P.N. Pathirana, A. Seneviratne, Blockchain and AI-based solutions to combat coronavirus (COVID-19)-like epidemics: A survey, IEEE Access (ISSN: 21693536) 9 (2021) 95730–95753, http://dx.doi.org/10.1109/ACCESS.2021.3093633.

[73] Q.-V. Pham, D.C. Nguyen, T. Huynh-The, W.-J. Hwang, P.N. Pathirana, Artificial intelligence (AI) and big data for coronavirus (COVID-19) pandemic: A survey on the state-of-the-arts, IEEE Access 8 (2021) 130820–130839, http://dx.doi.org/10.1109/ACCESS.2020.3009328.

[74] I. Castiglioni, L. Rundo, M. Codari, G. Di Leo, C. Salvatore, M. Interlenghi, F. Gallivanone, A. Cozzi, N.C. D'Amico, F. Sardanelli, AI applications to medical images: From machine learning to deep learning, Phys. Medica (ISSN: 1120-1797) 83 (2021) 9–24, http://dx.doi.org/10.1016/J.EJMP.2021.02.006.

[75] A. Self, Q. Chen, B.K. Desiraju, S. Dhariwal, A.D. Gleed, D. Mishra, R. Thiruvengadam, V. Chandramohan, R. Craik, E. Wilden, A. Khurana, S. Bhatnagar, A.T. Papageorghiou, J.A. Noble, Developing clinical artificial intelligence for obstetric ultrasound to improve access in underserved regions: Protocol for a computer-assisted low-cost point-of-care UltraSound (CALOPUS) study, J. MIR Res. Protoc. (ISSN: 1929-0748) 11 (9) (2022) e37374, http://dx.doi.org/10.2196/37374.

[76] F.-D. Lebit, P. DrRadu, L. Florentina Daniela, The role of 4D ultrasound in the assessment of fetal behaviour, Mædica (ISSN: 2069-6116) 6 (2) (2011) 120.

[77] M. Gottlieb, D. Nakitende, T. Sundaram, A. Serici, S. Shah, J. Bailitz, Comparison of static versus dynamic ultrasound for the detection of endotracheal intubation, West. J. Emerg. Med. (ISSN: 1936-9018) 19 (2) (2018) 412–416, http://dx.doi.org/10.5811/WESTJEM.2017.12.36714.

[78] N.K. Kerr, R. Dunham, S. Wolstenhulme, J. Wilson, Comparison of two- and three-dimensional transvaginal ultrasound in the visualisation of intrauterine devices, Ultrasound: J. Br. Med. Ultrasound Soc. (ISSN: 17431344) 22 (3) (2014) 141, http://dx.doi.org/10.1177/1742271X14532082.

[79] R. Chaoui, A. Abuhamad, J. Martins, K.S. Heling, Recent development in three and four dimension fetal echocardiography, Fetal Diagn. Ther. (ISSN: 1015-3837) 47 (2020) 345–353, http://dx.doi.org/10.1159/000500454.

[80] N.J. Magioli Bravo-Valenzuela, A.S. Malho, C.d. Nieblas, P.T. Castro, H. Werner, E. Araujo Jr., Evolution of fetal cardiac imaging over the last 20 years, Diagnostics (ISSN: 2075-4418) 13 (23) (2023) http://dx.doi.org/10.3390/diagnostics13233509.

[81] C.K. Abbey, M.A. Lago, M.P. Eckstein, Comparative Observer Effects in 2D and 3D Localization Tasks, vol. 8, (4) SPIE, (ISSN: 2329-4302) 2021, 041206, http://dx.doi.org/10.1117/1.JMI.8.4.041206.

[82] J. Yu, B. Yang, J. Wang, J.K. Leader, D.O. Wilson, J. Pu, 2D CNN versus 3D CNN for false-positive reduction in lung cancer screening, vol. 7, (5) SPIE, (ISSN: 2329-4302) 2020, 051202, http://dx.doi.org/10.1117/1.JMI.7.5.051202.

[83] E. Skelton, J. Matthew, Y. Li, B. Khanal, J.J. Cerrolaza Martinez, N. Toussaint, C. Gupta, C. Knight, B. Kainz, J.V. Hajnal, M. Rutherford, Towards automated extraction of 2D standard fetal head planes from 3D ultrasound acquisitions: A clinical evaluation and quality assessment comparison, Radiogr. (London, Engl. : 1995) (ISSN: 1532-2831) 27 (2) (2021) 519–526, http://dx.doi.org/10.1016/J.RADI.2020.11.006.

[84] M.B. Priya, C. Ramakrishnan, S. Karthik, Fetal 3D-echo classification and segmentation using color and textural features for TR detection, Revue Roumaine Sci. Techn. — Série Électrotechn. Énergétique (ISSN: 0035-4066) 69 (2024) 115–120, http://dx.doi.org/10.59277/RRST-EE.2024.1.19.

[85] J. Chen, Y. Zhang, J. Wang, X. Zhou, Y. He, T. Zhang, EllipseNet: Anchor-free ellipse detection for automatic cardiac biometrics in fetal echocardiography, in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 12907 LNCS, Springer Science and Business Media Deutschland GmbH, (ISSN: 16113349) ISBN: 9783030872335, 2021, pp. 218–227, http://dx.doi.org/10.1007/978-3-030-87234-2{_}21.

[86] J.N. Karim, E. Bradburn, N. Roberts, A.T. Papageorghiou, A.T. Papageorghiou, Z. Alfirevic, T. Chudleigh, H. Goodman, C. Ioannou, H. Longworth, J.N. Karim, K.H. Nicolaides, P. Pandya, G. Smith, B. Thilaganathan, J. Thornton, O. Rivero-Arias, H. Campbell, E. Juszczak, L. Linsell, E. Wilson, L. Hinton, J. Fisher, E. Duff, A. Rhodes, G. Yaz, First-trimester ultrasound detection of fetal heart anomalies: systematic review and meta-analysis, Ultrasound Obstet. Gynecol. (ISSN: 1469-0705) 59 (1) (2022) 11–25, http://dx.doi.org/10.1002/UOG.23740.

[87] J.S. Carvalho, G. Moscoso, Y. Ville, First-trimester transabdominal fetal echocardiography, Lancet (London, England) (ISSN: 0140-6736) 351 (9108) (1998) 1023–1027, http://dx.doi.org/10.1016/S0140-6736(97)08406-7.

[88] G.J. Burton, E. Jauniaux, Development of the human placenta and fetal heart: Synergic or independent? Front. Phys. (ISSN: 1664042X) 9 (4) (2018) 373, http://dx.doi.org/10.3389/FPHYS.2018.00373.

[89] R. Stoean, D. Iliescu, C. Stoean, V. Ilie, C. Patru, M. Hotoleanu, R. Nagy, D. Ruican, R. Trocan, A. Marcu, M. Atencia, G. Joya, Deep learning for the detection of frames of interest in fetal heart assessment from first trimester ultrasound, in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 12861 LNCS, Springer Science and Business Media Deutschland GmbH, (ISSN: 16113349) ISBN: 9783030850296, 2021, pp. 3–14, http://dx.doi.org/10.1007/978-3-030-85030-2{_}1.