Towards deep observation: A systematic survey on artificial intelligence techniques to monitor fetus via Ultrasound Images

MAHMOOD ALZUBAIDI1, MARCO AGUS1, KHALID ALYAFEI2,3, KHALED A ALTHELAYA1, UZAIR SHAH1, ALAA A. ABD-ALRAZAQ1, MOHAMMED ANBAR3, MICHEL MAKHLOUF2, AND MOWAfA HOUSEH1

1 College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar; 2 Weil Cornell Medical College-Qatar; 3Sidra Medical and Research Center, Doha, Qatar; 4 National Advanced IPv6 Centre, University Sains Malaysia, Penang, Malaysia

Key words: fetal; fetus; ultrasound; artificial intelligence; deep learning; machine learning

Abstract

Objective: Developing innovative informatics approaches aimed to enhance fetal monitoring is a burgeoning field of study in reproductive medicine. Over the last decade, several reviews have been conducted regarding Artificial intelligence (AI) techniques to improve pregnancy outcomes. They are limited by focusing on specific data (e.g., electronic health records) or a mother’s care during pregnancy. This survey aims to explore how artificial intelligence (AI) can assist with fetal growth monitoring via Ultrasound (US) image.

Materials and methods: We performed a systematic survey guided by the Joanna Briggs Institute Reviewer Manual. We reported our findings using the guidelines for PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis). We conducted a comprehensive search of eight medical and computer science bibliographic databases, including PubMed, Embase, PsycINFO, ScienceDirect, IEEE explore, ACM Library, Google Scholar, and the Web of Science. We retrieved studies published between 2010 to 2021. The search terms were chosen based on the target intervention (i.e., AI and ultrasound images) and target population (i.e., fetal/fetus). Publications written in English that converged on AI techniques used to monitor fetal care during pregnancy regardless of target fetal disease or anatomy were included. Data extracted from studies were synthesized using a narrative approach.

Results: Out of 1269 retrieved studies, we included 107 distinct studies from queries that were relevant to the topic in the survey. We found that 2D ultrasound images were more popular (n = 88) than 3D and 4D ultrasound images (n = 19). Classification is the most commonly used method (n = 42), followed by segmentation (n = 31), classification integrated with segmentation (n = 16) and other miscellaneous such as object-detection, regression and reinforcement learning (n = 18). The most common areas that gained traction within the pregnancy domain were the fetus head (n = 43), then fetus body (n=31), fetus heart (n=13), fetus abdomen (n=10), and lastly the fetus face (n=10). In the most recent studies, deep learning techniques were primarily used (n = 81), followed by machine learning (n=16), artificial neural network (n=7), and reinforcement learning (n=2). We did not find studies that developed AI-based applications at any medical center.

Conclusions: Overall, we found that AI techniques played a crucial role in predicting particular fetal diseases and identifying fetus anatomy structures during pregnancy. More research is required to validate this technology from a physician's perspective, such as pilot studies and randomized controlled trials on AI and its applications in a hospital setting. Additional research is required to focus in how radiologists can utilize AI technique because when AI is integrated into the clinical process as a tool to assist clinicians, more accurate and reproducible radiological assessments can be made.

1. Introduction

1.1 Background

Artificial Intelligence (AI) is a broad discipline that aims to replicate the inherent intelligence shown by people via artificial methods [1]. Recently, AI techniques have been widely utilized in the medical sector [2]. Historically, AI techniques were standalone systems with no direct link to medical imaging. With the development of new technology, the idea of ‘joint decision-making’ between people and AI offers the potential of boosting high performance in the area of medical imaging [3].

Correspondence to: Dr. Mowaf Househ, College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar
Tel.: +xxxxxxxxxxxxxxxxxxxxx; E-mail: mhouseh@hbku.edu.qa
In computer science, Machine Learning (ML), Deep Learning (DL), Artificial Neural Network (ANN) and Reinforcement Learning (RL) are subset techniques of AI that are used to perform different tasks on medical images such as classification, segmentation, object identification, and regression [4]–[6]. Diagnosis using computer-aided detection (CAD) has moved towards becoming AI automated process in the medical images[7], which include most of the medical imaging data such as (X-ray radiography, Fluoroscopy, MRI, Medical ultrasonography or ultrasound, Endoscopy, Elastography, Tactile imaging, and Thermography) [8]. However, digitized medical images come with a plethora of new information, possibilities, and challenges. Therefore, AI techniques are able to address some of these challenges by showing impressive accuracy and sensitivity in identifying imaging abnormalities. These techniques promise to enhance tissue-based detection and characterization, with the potential to improve diagnoses of diseases [9].

At present, the use of AI techniques in medical images has been discussed in depth across many medical disciplines, including identifying cardiovascular abnormalities, detecting fractures and other musculoskeletal injuries, aiding in diagnosing neurological diseases, reducing thoracic complications and conditions, screening for common cancers, and many other prognoses and diagnosis tasks [7], [10]–[13]. Furthermore, AI techniques have shown the ability to provide promising findings when utilizing prenatal medical images, such as monitoring fetal development at each stage of pregnancy, predicting a pregnant placenta’s health, and identifying potential complications [14]–[17]. AI techniques may assist with detecting several fetal diseases and adverse pregnancy outcomes with complex etiology and pathogenesis such as amniotic band syndrome, congenital diaphragmatic hernia, congenital high airway obstruction syndrome, fetal bowel obstruction, gastrochisis, omphalocele, pulmonary sequestration, and sacrococcygeal teratoma [18]–[20]. Therefore, more research is needed to understand the role of utilizing AI techniques in the early stage of pregnancy both to prevent and reduce unfavorable outcomes as well as provide an understanding of fetal anomalies and illnesses throughout pregnancy and for drastically reducing the need of more invasive diagnostic procedures that may be harmful for the fetus [21].

Various imaging modalities (e.g., Ultrasound, MRI, and computed tomography (CT)) are available and can be utilized by AI technique during pregnancy. [6] In medical literature, ultrasound imaging has become popular and is used during all pregnancy trimesters. Ultrasound is crucial for making diagnoses, along with tracking fetal growth and development. Additionally, Ultrasound can provide both precise fetal anatomical information as well as high-quality photos and increased diagnostic accuracy [17], [22]. There are numerous benefits and few limitations when using Ultrasound. Acquiring the device is a very inexpensive process, particularly when compared to other instruments used on the same organs, such as MRI or Positron Emission Tomography (PET). Unlike other acquisition equipment for comparable purposes, Ultrasound scanners are portable and long-lasting. The second major benefit is that, unlike MRI and CT, the Ultrasound machine does not pose any health concerns for pregnant women as the produced signals are entirely safe for both the mother and the fetus [23]. Although the literature exploring the benefits of utilizing AI technologies for Ultrasound imaging diagnostics in prenatal care has grown, more research is needed to understand the different roles AI can have in diagnostic prenatal care using ultrasound images.

1.2 Research Objective

Previous reviews on Ultrasound medical images for pregnant women using AI techniques has not been thoroughly conducted. Early reviews have attempted to summarize the use of AI in the prenatal period [22], [24]–[32]. However, these reviews are not comprehensive because they only target specific issues such as fetal cardiac function [24], Down syndrome [25], and the fetal central nervous system [26]. Similarly, other reviews have not focused on AI techniques and Ultrasound image as the primary intervention [27]–[29]. Two reviews covered the use of general AI techniques utilized in the prenatal period through Ultrasound. However, they did not focus on fetal as the primary population in their review [22], [30]. Another systematic review [31] addressed AI classification technologies for fetal health, but it only targeted cardiocriograpy disease. Other review [32] they introduced various areas of medical AI research including obstetrics but the main focus on US imaging in general. Therefore, a comprehensive survey is necessary to understand the role of AI technologies for the entire prenatal care period using Ultrasound images. This work will be the first comprehensive study that systematically reviews how AI techniques can assist during pregnancy screenings of fetal development and improve fetus growth monitoring. In this survey we aim to (a) Describe the application of AI techniques on different fetal spatial ultrasound dimensions (b) Discuss how different methodologies are used; classification, segmentation, object-detection, regression, and reinforcement learning, (c) Highlight the dataset acquisition and availability (d) Identify practical and research Implications and gaps for future research.

2. Methods

2.1 Overview

We conducted a survey using method of systematic reviews as reported by Joanna Briggs Institute for systematic reviews [33]. The findings are reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [34].

2.2 Search strategy

The bibliographic databases used in this study were PubMed, EMBASE, PsycINFO, IEEE Xplore, ACM Digital Library, ScienceDirect, and Web of Science. Google Scholar was also used as a search engine. The first 200 results were filtered from Google Scholar, which returned a significant number of papers sorted by relevancy to the search subject (20 pages). On June 22, 2021, the search began and ended on June 23, 2021. The original query was broad: we searched terms in all fields in all databases; due to the huge number of irrelevant studies while searching in Web of Science and ScienceDirect, we only searched article title, abstract, and keywords. We derived our preliminary search terms from previous works [22], [30] and identified additional terms in consultation with one digital health expert. Multimedia Appendix 1 lists the search strings used to search each bibliographic database.

2.3 Study Eligibility Criteria

Studies published in the last eleven years reporting on ‘fetus’ or ‘fetal’ were included in this survey. The population of interest was pregnant women in the first, second, and third trimesters with a specific focus on fetal development during these time periods. We included computer vision Artificial Intelligence interventions that
screened the fetal development process during the prenatal period. Ultrasound image was the only medical image addressed in this survey with different synonyms (e.g., Sonography, Ultrasonography, Echography, and Echocardiography). Inclusion and exclusion criteria are clearly defined in Table 1.

Table 1. Inclusion and exclusion criteria

| Criteria | Specified Criteria |
|----------|-------------------|
| **Inclusion** | • Studies that address fetus/fetal health during pregnancy that are involved AI and US images. |
| | • Studies published in English. |
| | • Peer-reviewed studies, conference proceedings, and book chapters. |
| | • Studies introducing new AI technology that is used for medical image processing. |
| | • Proposed algorithms that mainly perform the following tasks: classification, segmentation, object detection, regression, or reinforcement learning. |
| **Exclusion** | • Studies conducted from 2010 to 2021. |
| | • Studies that were out of the scope of fetal health. |
| | • Studies that do not use ultrasound as the source medical images such as MRI or CT. |
| | • Studies that relied on biometric ultrasound data instead of the images themselves. |
| | • Studies reported in a language other than English. |
| | • Conference abstracts, reviews magazines and newspapers. |
| | • Studies that use traditional image processing technologies without AI. |

**2.4 Study Selection**

A web-based systematic review tool called Rayyan Software aided the research selection process and sped up the screening step [35]. The studies were chosen using a two-step process. First, all the retrieved papers’ titles and abstracts were independently checked by two authors. Both authors independently read the entire texts of the studies considered in the first phase. Any disagreement in the screening process was resolved through consensus by the authors involved.

**2.5 Data Extraction and Synthesis**

To undertake a systematic data extraction, we designed a data extraction form and piloted it using three included studies (Multimedia Appendix 2). Data from the included studies was extracted separately by two authors. Discrepancies were addressed by consultation with the other authors. The data was synthesized using a narrative method and the results of the research were categorized into groups based on Fetal organ and AI technique tasks.

**3. Results**

**3.1 Search Results**

As shown in Figure 1, the search yielded a total of 1269 citations. After excluding 457 duplicates, there were 812 unique titles and abstracts. A total of 598 citations were excluded after evaluating the titles and abstracts. After full-text screening, n= 107 citations were excluded from the remaining n= 214 papers. The narrative synthesis includes a total of 107 studies. Across screening steps, both authors reported the excluding reason for each study to ensure reliability of the work. The total excluded studies were n= 704 articles. These studies did not meet selection criteria due to the following reasons: (1) irrelevant, (2) wrong intervention, (3) wrong population, (4) wrong publication type, (5) unavailable, and (6) foreign language article. These terminologies are defined in Table 2.

**Table 2. Definition of the excluded terminologies**

| Terminologies | Definition |
|---------------|------------|
| Irrelevant    | Publication that are not related to the scope of this survey |
| Wrong intervention | Publication that targets fetus health but not using AI technology and Ultrasound image |
| Wrong population | Publication that use AI technology and Ultrasound image but did not target fetus health |
| Wrong publication type | Publication that is conference abstract, review, magazine, and newspaper |
| Unavailable   | Publication that is not accessible or cannot be found |
| Foreign language | Publication that is not written in English language |

**3.2 Assessment of Bibliometrics Analysis**

To validate the selected studies, we used a bibliometrics analysis [36] tool that explores the literature through citations and automatically recommends highly related articles to our scope. This ensured that we were unlikely to miss any relevant study. Figure 2. Provides a citations map between selected studies over time. The map shows that all selected studies are relevant over time from 2010 to 2021 and recommends nine studies that may be relevant to our scope based on the interactive citation. However, these studies were excluded because they did not meet our inclusion criteria. Therefore, we concluded that our search was comprehensive and included most of the relevant studies.

Figure 2. Literature Map showing the selected studies in red dot and recommended studies in black dot.
3.3 Identification of result themes

In Figure 1, we grouped the studies by the fetus organ that each study addresses including: 1) fetus head (n=43, 40.18%), 2) fetus body (n=31, 28.97%), 3) fetus heart (n=13, 12.14%), 4) fetus face (n=10, 9.34%), and 5) fetus abdomen (n=10, 9.34%). Each group is classified by subgroup as detailed in Figure 3. In addition, most of the studies used 2D US (n=88, 82.24%) and 3D/4D US were rare and reported only in (n=19, 17.75%). AI tasks are grouped based on the most common use. We found that classification was used in (n=42, 39.25%), segmentation was used in (n=31, 28.97%), both classification and segmentation were used together in (n=16, 14.95%), object-detection, regression, and reinforcement learning were seen as miscellaneous in (n=18, 16.82%).

3.4 Definition of result themes

3.4.1 Ultrasound imaging modalities

There was a total of (n=88, 82.24%) studies that utilized 2D US images. Therefore, here is a brief about what’s the 2D Ultrasound. Sound waves are used in all ultrasounds to produce an image. The conventional ultrasound picture of a growing fetus is a 2D image. The fetus’s internal organs may be seen with a 2D ultrasound, which provides outlines and flat-looking pictures. 2D ultrasounds have been widely available for a long time and have a very good safety record. These devices do not pose the same dangers as X-rays because they employ non-ionizing rather than ionizing radiation [22]. Ultrasounds are typically conducted at least once throughout pregnancy, most often between 18 and 22 weeks in the second trimester [37]. This examination, also known as a level two ultrasound or anatomy scan, is performed to monitor fetus’s development. Ultrasounds may be used to examine a variety of things during pregnancy, including: 1) how the fetus is developing; 2) the gestational age of the fetus; 3) any abnormalities within the uterus, ovaries, cervix, or placenta; 4) The number of fetuses you’re carrying; 5) Any difficulties the fetus may be experiencing; 6) the fetus’s heart rate; 7) fetal growth and location in the uterus, 8) amniotic fluid level, 9) sex assignment, 10) signs of congenital defects, and 11) signs of Down syndrome [38], [39].

There was a total of (n=18, 16.8%) studies that utilized 3D. Therefore, here is a brief about what’s the 3D Ultrasound. Observing three-dimensional (3D) ultrasound images has increasingly grown in recent years. Although 3D ultrasounds may be beneficial in identifying facial or skeletal abnormalities [40], 2D ultrasounds are commonly utilized in medical settings because they can clearly reveal the interior organs of a growing fetus. The picture of a 3D ultrasound is created by putting together several 2D photos obtained from various angles. Many parents like 3D photos because they believe they can see their baby’s face more clearly than they can with flat 2D images. Despite this, the Food and Drug Administration (FDA) does not recommend using a 3D ultrasound for entertainment purposes [41]. The “as low as reasonably achievable” (ALARA) approach directs ultrasound technologists to be used solely in a clinical setting in order to reduce exposure to heat and radiation [42]. While ultrasound is generally thought to be harmless, there is insufficient data to determine what long-term exposure to ultrasound may do to a fetus or a pregnant woman. There is no way of knowing how long a session will take or if the ultrasound equipment will work correctly in non-clinical situations such as places that give “keepsake” photos [43].

4D ultrasound is identical to a 3D ultrasound, but the image it creates is updated constantly - comparable to a moving image. This sort of ultrasound is usually performed for fun rather than for medical concerns. As previously mentioned, because ultrasound is medical equipment that should only be used for medical purposes [44], the FDA does not advocate getting ultrasounds for enjoyment or bonding purposes. Unless their doctor or midwife has recommended it as part of prenatal care, pregnant patients should avoid non-medical locations that provide ultrasounds[43]. Though there’s no proven risk but both 3D and 4D ultrasounds employ higher-than-normal amounts of ultrasound energy and sessions are longer compared to 2D which may have fetal adverse effects[45], [46].

3.4.2 AI image processing task

There was a total of (n=42, 39.25%) studies that utilized classification, which is the process of assigning one or more labels to an image, and it is one of the most fundamental problems concerning accurate computer vision and pattern recognition. It has many applications, including image and video retrieval, video surveillance, web content analysis, human-computer interaction, and biometrics. Feature coding is an important part of image classification, and several coding methods have been developed in recent years. In general, image classification entails extracting picture characteristics before classifying them. As a result, the main aspect of utilizing image classification is understanding how to extract and evaluate image characteristics [47].

There was a total of (n=31, 28.97%) studies that utilized
segmentation, which one of the most complex task in medical image processing, that distinguishing the pixels of organs or lesions from background medical images such as CT or MRI scans. A number of researchers have suggested different automatic segmentation methods by using existing technology[48]. Previously, traditional techniques like edge detection filters and mathematical algorithms were used to construct earlier systems [49]. After this, machine learning techniques for extracting hand-crafted characteristics were the dominating method for a substantial time period. The main issue for creating such a system has always been designing and extracting these characteristics, and the complexity of these methods have been seen as a substantial barrier to deployment. In the 2000’s, deep learning methods began to grow in popularity because of advancements in technology as this method had significantly better skills in image processing jobs. Deep learning methods have emerged as a top choice for image segmentation, particularly medical image segmentation, due to their promising capabilities [50]. There was a total of (n=12, 11.21%) studies that utilized object detection techniques, which is the process of locating and classifying items. A detection method is used in biomedical images to determine the regions where the patient’s lesions are situated as box coordinates. There are two kinds of deep learning-based object detection. These region proposal-based algorithms are one example. Using a selective search technique, this method extracts different kinds of patches from input images. After that, the trained model determines if each patch contains numerous items and classifies them according to their region of interest (ROI). In particular, the region proposal network, was created to speed up the detection process. Object identification in the other methods is done using the regression-based algorithm as a one-stage network. These methods use image pixels to directly identify and detect bounding box coordinates and class probabilities within entire images [6].

There was a total of (n=2, 1.8%) studies that utilized Reinforcement Learning (RL), the concept behind RL is that an artificial agent learns by interacting with its surroundings. It enables agents to autonomously identify the best behavior to exhibit in each situation to optimize performance on specified metrics. The basic concept underlying reinforcement learning is made up of many components. The RL agent is the process’s decision-maker, and it tries to perform an action that has been recognized by the environment. Depending on the activity performed, the agent receives a reward or penalty from its surroundings. Exploration and exploitation are used by RL agents to determine which activities provide the most rewards. In addition, the agent obtains information about the condition of the environment [51]. For medical image analysis applications, RL provides a powerful framework. RL has been effectively utilized in a variety of applications, including landmark localization, object detection, and registration using image-based parameter inference. RL has also been shown to be a viable option for handling complex optimization problems such as parameter tuning, augmentation strategy selection, and neural architecture search. Existing RL applications for medical imaging can be split into three categories: parametric medical image analysis, medical image optimization, and numerous miscellaneous applications [52].

3.5 Description of Included Studies
As shown in Table 3, more than half of the studies (n= 53, 49.53%) were obtained from journal articles, while the other half were found in conference proceedings (n= 43, 40.18%) and book chapters (n=11, 10.28%). Studies were published between 2010 and 2021. The majority of studies were published in 2020 (n=30, 28.03%), followed by 2021 (n=19, 17.75%), 2019 (n=14, 13.00%), 2018 (n=11, 10.28), 2014 (n=9, 8.41%), 2017 (n=6, 5.60%), 2015 (n=6, 5.60%), 2016 (n=4, 3.73%), 2013 (n=3, 2.80%), 2011 (n=3, 2.80%), 2012 (n=1, 0.93%), and 2010 (n=1, 0.98%). In regard to studies examining fetal organs, fetus skull localization and measurement was reported on in (n=25, 23.36%) studies, followed by fetal part structures (n=13, 12.14%) and brain standard plane in (n=13, 12.14%) studies. Abdominal anatomical landmarks were reported in (n=10, 9.34%) studies. Additionally, fetus body anatomical structures were reported in (n=8, 7.47%) studies. Heart disease was reported on in (n=7, 6.54%) studies. Growth disease (n=6, 5.60%) and brain disease were also reported in (n=5, 4.67%) studies. The view of heart chambers was reported on in (n=6, 5.60%) studies. Fetal facial standard planes were reported in (n=5, 4.67%) studies. Gestational age (n=3, 2.80%) and face anatomical landmarks were reported on in (n=3, 2.80%) studies. Facial expressions were reported on in (n=2, 1.86%) studies, and gender identification was reported on only in (n=1, 0.93%) study. In addition, most of the studies were conducted in China (n=41, 38.31%), followed by the UK (n=25, 23.36%), and India (n=14, 13.08%). Surprisingly, some institutes mainly focused on this field and contributed high number of studies. For example, many (n=20, 18.69%) studies were conducted at the University of Oxford, UK and several (n=13, 12.14%) were conducted in Shenzhen University, China.

3.6 Organization of the survey findings
The survey is organized as shown in Figure 3. We divided AI for Fetal US medical image into five parts (Section 4), which classify fetus research by organs. In the US image, each group was divided into sub-sections based on recognized characteristics or target illnesses. This survey discusses how different tasks are used; classification, segmentation, object-detection, regression, and reinforcement learning. We discuss the leverage of public dataset in Section 5 and compared similar work depending on the fetal organ, as shown in Section 4. We also divided the discussion (Section 6) into two subsections: principle findings, and practical and scientific implications. Lastly, we highlighted research gaps and future work in Section 7.
### Table 3. Characteristics of the included studies (n=107).

| Characteristics                                      | Number of studies |
|------------------------------------------------------|-------------------|
| **Type of publication**                              | (n, %)            |
| Journal article                                      | (53, 49.53)       |
| Conference proceedings                               | (43, 40.18)       |
| Book chapter                                         | (11, 10.28)       |
| **Year of publication**                              |                   |
| 2021                                                 | (19, 17.75)       |
| 2020                                                 | (30, 28.03)       |
| 2019                                                 | (14, 13.08)       |
| 2018                                                 | (11, 10.28)       |
| 2017                                                 | (6, 5.60)         |
| 2016                                                 | (4, 3.73)         |
| 2015                                                 | (6, 5.60)         |
| 2014                                                 | (9, 8.41)         |
| 2013                                                 | (3, 2.80)         |
| 2012                                                 | (1, 0.93)         |
| 2011                                                 | (3, 2.80)         |
| 2010                                                 | (1, 0.93)         |
| **Fetal Organ**                                      |                   |
| Fetus Body                                           |                   |
| Fetal part structures                                | (31, 28.97)       |
| Anatomical structures                                | (13, 12.14)       |
| Growth disease                                       | (8, 7.47)         |
| Gestational age                                      | (6, 5.60)         |
| Gender                                               | (3, 2.80)         |
| Fetus Head                                           | (43, 40.18)       |
| Skull localization and measurement                   | (25, 23.36)       |
| Brain standard plane                                 | (13, 12.14)       |
| Brain disease                                        | (5, 4.67)         |
| Fetus Face                                           | (10, 9.34)        |
| Fetal facial standard planes                         | (5, 4.67)         |
| Face anatomical landmarks                            | (3, 2.80)         |
| Facial expressions                                   | (2, 1.86)         |
| Fetus Heart                                          | (13, 12.14)       |
| Heart disease                                        | (7, 6.54)         |
| Heart chambers view                                  | (6, 5.60)         |
| **Fetus Abdomen**                                    | (10, 9.34)        |
| Abdominal anatomical landmarks                       | (10, 9.34)        |
| **Publication Country and top institute**            |                   |
| China                                                | (41, 38.31)       |
| Shenzhen University                                  | (13, 12.14)       |
| Beihang University                                  | (5, 4.67)         |
| Chinese University of Hong Kong                     | (4, 3.73)         |
| Hebei University of Technology                       | (2, 1.86)         |
| Fudan University                                    | (2, 1.86)         |
| Shanghai Jiao Tong University                       | (2, 1.86)         |
| South China University of Technology                 | (2, 1.86)         |
| Other institutes                                     | (11, 10.28)       |
| UK                                                   | (25, 23.36)       |
| University of Oxford                                | (20, 18.69)       |
| Imperial College London                             | (4, 3.73)         |
| King’s College, London                              | (1, 0.93)         |
| India                                                | (14, 13.08)       |
| Japan                                                | (4, 3.73)         |
| Indonesia                                           | (4, 3.73)         |
| USA                                                  | (3, 2.80)         |
| South Korea                                         | (3, 2.80)         |
| Iran                                                 | (3, 2.80)         |
| Australia                                           | (1, 0.93)         |
| Canada                                               | (1, 0.93)         |
| Mexico                                               | (1, 0.93)         |
| France                                               | (1, 0.93)         |
| Italy                                                | (1, 0.93)         |
| Tunisia                                              | (1, 0.93)         |
| Spain                                                | (1, 0.93)         |
| Iraq                                                 | (1, 0.93)         |
| Brazil                                               | (1, 0.93)         |
| Malaysia                                             | (1, 0.93)         |
Figure 3. Organization of this survey paper
4. Artificial intelligence for fetus Ultrasound

4.1 Fetus Body

As shown in Table 3, the overall purpose of several (n=31, 28.97%) studies was to identify general characters of the fetus (e.g., gestational age, gender) or diseases such as Intrauterine Growth Restriction (IUGR). In addition, most of the studies aimed to identify the fetus itself in the uterus or fetus part structure during different trimesters. The first trimester is from week 1 to the end of week 12. The second trimester is from week 13 to the end of week 26, and the third trimester is from week 27 to the end of the pregnancy. We also report on a performance comparison between the various techniques for each fetus body group including objective, backbone methods, optimization, fetal age, best obtained result, and observations (see Table 5).

4.1.1 Fetal Part Structures

- **Classification**

  For identification and localization of fetal part structure, image classification was widely used in (n=9, 8.41%) studies. Only 2D US images were utilized for classification purposes. Studies [53]–[55] used a classification task to identify and locate fetal skull, heart, and abdomen from 2D US images in the second trimester. Besides that, in [56], classification was used to locate the exact abdominal circumference plane (ACP), and in [57], more results were obtained by locating head circumference plane (HCP), the abdominal circumference plane (ACP), and the femur length plane (FLP). Furthermore, in studies [58]–[60] researching the second and third trimester, multiclasses were used to identify and locate various parts including fetal abdominal standard plane (FASP); 2) fetal face axial standard plane (FFASP); and 3) fetal four-chamber view standard plane (FFVSP) of heart. Study [61] located the mother’s cervix in addition to abdomen, brain, femur, and thorax.

- **Segmentation**

  Image segmentation task was used in (n=3, 2.80%) studies for the purpose of fetal part structure identification. 2D US imaging [62] was used to segment neonatal hip bone including seven key structures: 1) chondro-osseous border (CB), 2) femoral head (FH), 3) synovial fold (SF), 4) joint capsule & perichondrium (JCP), 5) labrum (La), 6) cartilaginous roof (CR), and 7) bony roof (BR). Only one study in [63] provided segmentation model that able to segment fetus kidney using 3D US in the third trimester. Segmentation was used to locate fetal head, femur, and humerus using 2D US as seen in [64].

- **Classification and segmentation**

  In [65], whole fetal segmentation followed by classification tasks were used to locate the head, abdomen (in sagittal view), and limbs (in axial view) using 3D US images taken in the first trimester.

4.1.2 Anatomical Part Structures

- **Classification**

  For identification and localization of anatomical structures, image classification was used in (n=4, 3.73%) studies. Only one study [66] used 3D US images and other studies used 2D US [67]–[69]. In [66], the whole fetus was localized in the sagittal plane and classifier was then applied to the axial images to localize one of three classes (head, body, and non-fetal) during the second trimester. Multiclassification methods were utilized in [67] to localize head, spine, thorax, abdomen, limbs, and placenta in the third trimester. Moreover, binary classification was used in [68] to identify abdomen versus non-abdomen. In [69], classification method was used to detection and measurement nuchal translucency (NT) in the beginning of second trimester. Moreover, monitoring NT combined with maternal age can provide effective insight of screening Down syndrome.

  - **Segmentation**

    An image segmentation task was used in (n=4, 3.73%) studies for the purpose of image segmentation was target task anatomical structures identification. In [70], [71], 3D US were utilized to segment fetus, gestational sac amniotic fluid, and placenta in the beginning of second trimester. However, in [72], 2D US was used to segment amniotic fluid and the fetus in the late trimester. Finally, 2D US in [73] was utilized to segment and measure nuchal translucency (NT) in the first trimester of pregnancy.

4.1.3 Growth disease diagnosis

- **Classification**

  For diagnosis of fetal growth, disease image classification was conducted in (n=4, 3.73%) studies using 2D US. In [74] binary classification was used for early diagnosis of Intrauterine Growth Restriction (IUGR) at the third trimester of pregnancy. The features considered to determine a diagnosis of IUGR are gestational age (GA), biparietal diameter (BPD), abdominal circumference (AC), head circumference (HC), and femur length (FL). Additionally, binary-classification was used in the second trimester of pregnancy to identify normal vs abnormal fetus growth as seen in [75]. In [76], binary-classification is used to identify the normal growth of the ovum by distinguishing Blighted Ovum from healthy Ovum. Lastly, the binary classification was used to distinguish between healthy fetuses and those with Down syndrome [77].

- **Classification and segmentation**

  For diagnosis of fetal growth disease, image classification along with segmentation were used in (n=2, 1.86%) studies using 2D US. In [78], [79], segmentation of the region of interest (ROI), followed by classification to diagnosis (IUGR) (normal vs abnormal) were carried out. This was done by using US Images taken in both second and third trimester of pregnancy. This classification is used to estimate the measurement of the following: 1) Fetal abdominal circumference (AC), 2) Head circumference (HC), 3) Biparietal diameter (BPD), 4) Femur length.

4.1.4 Gestational age (GA) estimation

- **Classification and segmentation**

  Both classification as well as segmentation tasks were used to estimate GA in (n=3, 2.80%) studies using 2D US taken in the second and third trimester of pregnancy. In [80], the trans-cerebellar diameter (TCD) measurement was used to estimate GA in week. The TC plane frames are extracted from the US images using classification. Segmentation then localizes the TC structure and performs automated TCD estimation, from which the GA can thereafter be estimated via an equation. In [81], 2D US was used to estimate GA based on region of fetal lung in the second and third trimester. In the first stage, the segmentation is to learn the recognition of fetal lung region in the ultrasound images. Classification is also used to accurately estimate the gestational age from the fetal lung region of ultrasound images. Several fetal structures were utilized to estimate GA [82], which focused on the second trimester. In this study, 2D US images were classified into four categories: head (BPD and HC), abdomen (AC), femur (FL), and fetus (crown-rump length: CRL). Then, the regions of interest (i.e., head, abdomen, and femur) were segmented and the results of biometry measurements were used to estimate GA.

4.1.5 Gender identification

- **Classification**

  Binary-classification was used in [83] to identify the
4.2 Fetus Head

As shown in Table 3, the primary purpose of (n=43, 40.18%) studies is to identify and localize fetus skull (e.g., head circumference), brain standard planes (e.g., Lateral sulcus (LS), thalamus (T), choroid plexus (CP), cavum septi pellucidi (CSP)) or brain diseases (e.g., hydrocephalus (premature GA), ventriculomegaly, central nervous system (CNS) malformations). The following subsection discusses each category based on the implemented task. Table 6 presents comparison between the various techniques for each fetus head group including objective, backbone methods, optimization, fetal age, best obtained result, and observations. studies for each group under the fetus head category and provides an overview of methodology and observations.

4.2.1 Skull localization and measurement

- **Classification**
  
  Classification tasks for skull localization and head circumference are rarely used, as seen only in (n=3, 2.80%) studies. Studies [84], [85] used classification to localize the region of interest (ROI) and identify the fetus head based on 2D US taken in first and second trimester respectively. 3D US taken in first trimester were used to detect a fetal head in one study [86].

- **Segmentation**
  
  Skull localization and head circumference (HC) in most of the studies (n=16, 14.81%) used segmentation task. 2D US was used in 16 studies and one study used 3D US. Various network architecture were used to segment and locate skull, then perform head circumference as seen in [87]. Besides identifying HC, in 100%segmentation was also used to find fetal biparietal diameter (BPD). Further investigation shows that the work completed in [101] was the first investigation about whole fetal head segmentation in 3D US. The segmentation network for skull localization in [91] was tested to identify a view of the four heart chambers on different datasets.

- **Miscellaneous**
  
  Along with segmentation, another method was used to locate and identify the fetus skull in (n=6, 5.60%) studies. As seen in [102], regression was used along with segmentation to identify HC, BPD, and occipitofrontal diameter (OFD). Besides this, in [103], neural network was used to train the saliency predictor and predict saliency maps to measure HC and identify the alignment of both trans ventricular (TV) as well as trans cerebellar (TC). In addition, an object detection task was used along with segmentation in [104] to locate the fetal head at the end of the first trimester. In [105], object detection was used for head localization and centering and regression to delineate the HC accurately. In [106], classification and segmentation tasks were used by utilizing a multi-Task network to identify head composition.

4.2.2 Brain standard plane

- **Classification**
  
  The classification was used in (n=2, 1.86%) studies to identify the brain standard plane, and both studies used 2D US images taken in the second and third trimesters. In [107], [108], two different classification network architectures were used to identify six fetal brain standard planes. These architectures include horizontal transverse section of the thalamus, horizontal transverse section of the lateral ventricle, transverse section of the cerebellum, midsagittal plane, paracentral sagittal section, and coronal section of the anterior horn of the lateral ventricles.

- **Segmentation**
  
  The segmentation task was used in (n=3, 2.80%) studies to identify the brain standard plane. Each of these studies used 2D US images. In [109], 2D US images taken in the second trimester were used to segment fetal cerebellum structure. In addition, in [110], 2D US images taken in the third trimester were used to segment fetal Middle Cerebral Artery (MCA). Authors in [111] utilized 3D US images in the second trimester to formulate the segmentation as a classification problem to identify the following brain planes; background, Choroid Plexus (CP), Lateral Posterior Ventricle Cavity (PVC), the Cavum Septum Pellucidi (CSP), and Cerebellum (CEB).

- **Classification and segmentation**
  
  Classification with segmentation is employed in (n=3, 2.80%) studies to identify the brain standard plane. These studies all used 3D US images. In [112], both tasks were used to identify brain alignment based on skull boundaries and then head segmentation, eye localization, and prediction of brain orientation in the second and third trimesters. Moreover, in [113], segmentation followed by a classification task was used to detect cavum septi pellucidi (CSP), thalami (Tha), lateral ventricles (LV), cerebellum (CE), and cisterna magna (CM) in the second trimester. Segmentation and classification are employed by 3D US in [114] to identify the skull, midsagittal plane, and orbits of the eyes in the second trimester of pregnancy.

- **Miscellaneous**
  
  For brain standard planes identification in (n=5, 4.62%), methods different than previously mentioned were used, including segmentation with object detection, object detection, classification with object detection, and Reinforcement Learning (RL). As seen in [115], the RL-based technique was employed for the first time on 3D US images to localize standard brain planes, including trans-thalamic (TT) and trans-cerebellar (TC) in the second and third trimesters. RL is used in [116] to localize the following: mid-sagittal (S), transverse (T), and coronal (C) planes in volumes and trans-thalamic (TT), trans-ventricular (TV), and trans-cerebellar (TC). Object detection architecture is utilized on 2D US images in [117] to localize the trans-thalamic plane in the second trimester of pregnancy, including lateral sulcus (LS), thalamus (T), choroid plexus (CP), cavum septi pellucidi (CSP), and third ventricle (TV). In [118], classification with object detection was used to detect Lateral sulcus (LS), thalamus (T), choroid plexus (CP), cavum septi pellucidi (CSP), third ventricle (TV), brain midline (BM). In contrast,[119]used classification with object detection to localize cavum septum pellucidum (CSP) and ambient cistern (AC) and cerebellum), as well as to measure head circumference (HC) and biparietal diameter (BPD).

4.2.3 Brain disease

- **Classification**
  
  The Binary-classification technique utilized 2D US images taken in the second trimester as seen in [120] to detect hydrocephalus disease using premature GA.

- **Classification and segmentation**
  
  Studies [121], [122] both used 2D US images taken in the second and third trimesters to segment craniocerebral and identify abnormalities or specific diseases. As seen in [121], central nervous system (CNS) malformations were detected using binary classification. Moreover, in [122], multiclassification was used to identify the following problems: TV planes contained occurrences of ventriculomegaly and hydrocephalus, and TC planes contained occurrences of Blake pouch cyst (BPC).

- **Miscellaneous**
In [123], [124] different methods were used to detect ventriculomegaly disease based on 2D US images taken in the second trimester. In study [123], object detection and regression first identify fetal brain ultrasound images from standard axial planes as normal or abnormal. Second regression was then used to find the lateral ventricular regions of images with big lateral ventricular width. Then, the width was anticipated with a modest error based on these regions. Furthermore, [124] was the first study to propose an object-detection-based automatic measurement approach for fetal lateral ventricles (LVs) based on 2D US images. The approach can both distinguish and locate the fetal LV automatically as well as measure the LV's scope quickly and accurately.

4.3 Fetus Face

As shown in Table 3, the primary purpose of (n= 10, 9.34%) studies was to identify and localize fetus face features such as the fetal facial standard plane (FFSP) (i.e., axial, coronal, and sagittal plane), face anatomical landmarks (e.g., Nasal Bone), and facial expressions (i.e., sad, happy). The following subsection discusses each category based on the implemented task. Table 7 presents comparison between the various objectives for each fetus face group including objective, backbone methods, optimization, fetal age, best obtained result, and observations.

4.3.1 Fetal facial standard plane (FFSP)

- **Classification**
  
  Classification was used to classify the FFSP using 2D US images as seen in (n=5, 4.67%). Using 2D US images that taken in the second and third trimesters, four studies [125]–[128] identify ocular axial planes (OAP), the median sagittal planes (MSP), and the nasolabial coronal planes (NCP). In addition, authors in [129] were able to identify the FFSP using 2D US images taken in the second trimester.

4.3.2 Face anatomical landmarks

- **Miscellaneous**
  
  Different methods were used to identify face anatomical landmarks in (n=3, 2.80%) studies. In [130], 3D US images taken in the second and third trimesters were segmented to identify background, face mask (excluding facial structures), eyes, nose, and lips. In another study the object detection method was used on 3D US images in [131] to detect the left fetal eye, middle eyebrow, right eye, nose, and chin. Furthermore, classification used 2D US images taken in the first and second trimesters to detect the nasal bone. This was done to enhance the detection rate of Down syndrome, as seen in [132].

4.3.3 Facial expressions

- **Classification**
  
  For the first time, 4D US images were utilized using the multi-classification method to identify fetal facial expression into Sad, Normal, and Happy as seen in [133]. Study [134] used 2D US images in the second and third trimesters to classify fetal facial expression into eye blinking, mouthing without any expression, scowling, and yawning.

4.4 Fetus Heart

As shown in Table 3, the primary purpose of (n= 13, 14.04%) studies is to identify and localize heart diseases and the fetus heart chambers view. The following subsection discusses each category based on the implemented task. Table 8 presents comparison between the various techniques for each fetus heart group including objective, backbone methods, optimization, fetal age, best obtained result, and observations.

4.4.1 Heart disease

- **Classification**
  
  Two studies used classification methods to identify heart diseases based on 2D US images taken in the second trimester. In [135], binary-classification task was utilized to identify Down syndrome vs normal fetuses based on identifying the echogenic intracardiac foci (EIF). Furthermore, in [136], binary-classification task was utilized to detect hypoplastic left heart syndrome (HLHS) vs healthy cases based on the four-chamber heart (4CH), left ventricular outflow tract (LVOT), and right ventricular outflow tract (RVOT).

- **Segmentation**
  
  2D US images were utilized by multi-class segmentation to identify heart disease, including: left heart syndrome (HLHS), total anomalous pulmonary venous connection (TAPVC), pulmonary atresia with intact ventricular septum (PA/IVS), endocardial cushion defect (ECD), fetal cardiac rhabdomyoma (FRC), and Ebstein's anomaly (EA) [137].

- **Classification and segmentation**
  
  Classification and segmentation are used to detect congenital heart disease (CHD) based on 2D US images taken in the second trimester [138]. Cases were classified into normal hearts vs CHD. This classification was orchestrated by identifying five views of the heart in fetal CHD screening, including three-vessel trachea (3VT), three-vessel view (3VV), left-ventricular outflow tract (LVOT), axial four chambers (A4C), and abdomen (ABDO).

- **Miscellaneous**
  
  To identify heart disease, various methods were proposed based on the capabilities of 2D US images. Classification with object detection were utilized in [139], [140]. Fetal congenital heart disease (FHD) can be detected in the second and third trimesters based on how quickly the fetus grows between gestational weeks, and how the shape of the Four Chambers heart changes over time [139]. Furthermore, classification with object detection was used to detect cardiac substructures and structural abnormalities in the second and third trimesters [140]. Lastly, segmentation with object detection was used to locate the ventricular septum in 2D US images in the second trimester, as seen in [141].

4.4.2 Heart chambers view

- **Classification**
  
  2D US images taken in the second and third trimesters were utilized in [142] to propose a classification task utilized to localize the four chambers (4CH), the left ventricular outflow tract (LVOT), the three vessels (3V), and the background (BG).

- **Segmentation**
  
  Segmentation was used in (n=3, 2.80%) studies based on 2D US images. In two studies [143], [144], seven critical anatomical structures in the apical four-chamber (A4C) view were segmented, including: left atrium (LA), right atrium (RA), left ventricle (LV), right ventricle (RV), descending aorta (DAO), epicardium (EP) and thorax. In another study [145], segmentation was used to locate four-chamber views (4C), left ventricular outflow tract view (LVOT), and three-vessel view (3V) in the second and third trimesters.

- **Miscellaneous**
  
  2D US images were used in (n=2, 1.86%) studies, and both studies utilized classification with object detection. In the second trimester [146], the first classification conducted was the cardiac four-chamber plane (CFF) into non-CFPs and CFPs (i.e., apical, bottom, and parasternal CFPs). The second task was then classify the CFPs in terms of the zoom and gain of 2D US images. In addition, object detection was utilized to detect anatomical structures in the CFPs, including left atrial pulmonary vein angle (PVA), apex cords and moderator band (ACMB), and multiple ribs (MRs). In [147], object detection was used in the second and third trimesters to extract attention regions for improving classification performance and determining the four-chamber view, including detection of end-systolic (ES) and end-diastolic (ED).

4.5 Fetus Abdomen

As shown in Table 3, the primary purpose of (n= 10,
10.8% (n=7, 6.54%) studies was to identify and localize the fetus’s abdomen, including abdominal anatomical landmarks (i.e., stomach bubble (SB), umbilical vein (UV), and spine (SP)). The following subsection discusses this category based on the implemented task. Table 9 presents comparison between the various techniques for each fetus abdomen including objective, backbone methods, optimization, fetal age, best obtained result, and observations.

4.5.1 Abdominal anatomical landmarks
- Classification
  Classification task was used in (n=7, 6.54%) studies to localize abdominal anatomical landmarks. 2D US images were used in seven studies [148]–[153] and 3D US images in only one study [154]. In [148], [150], [152], [154], various classifiers were utilized to localize stomach bubble (SB) and umbilical vein (UV), in the first and second trimester as seen in [148], [154], and in the second and third trimesters as seen in [150], [152]. Furthermore, works in [149], [151], [153] located the spine (SP) besides SB and UV on the same trimesters.
- Classification and segmentation
  2D US images taken within random trimesters were used in [155], [156]. In [155], SB, UV, and AF were localized. Further, abdominal circumference (AC), spine position, and bone regions were estimated. In [156], 2D US images were classified into normal vs abnormal fetuses based on the fetus images such as AF, SB, UV, and SA. Study [157] located SB and the portal section from the UV, and observed amniotic fluid (AF) in the second and third trimesters.

5. Dataset Analysis

5.1 Public dataset

5.1.1 Fetus Body
In this area, ethical and legal concerns presented the most barriers to comprehensive research. Therefore, datasets were not available to the research community. For example, in this research, studies discussed how the fetal body contains five subsections (fetal part structure, anatomical structure, growth disease, gestational age, and gender identification). Out of the available studies (n=31, 28.9%), one dataset was intended to be available online in [55] related to identifying fetal part structure. Unfortunately, this dataset has not yet been released by the author. The only public dataset that was available online for fetal structure classification was released by Burgos-Artizzu et al. [61], the dataset is a total of 12499 2D fetal ultrasound images, including brain 143, Trans-cerebellum 714, Trans-thalamic 1638, Trans-ventricular 597, abdomen 711, cervix 1626, femur 1040, thorax 1718, and 4213 are unclassified fetal image. Acquiring a high volume of the dataset was also challenging in most of the studies; therefore, multi-data augmentation in (n=11, 10.2%) studies [54], [56], [61], [63], [65], [70], [72], [76], [80] has been observed. These augmented images were employed to boost classification and segmentation performance. SimpleITK library [158] was used for augmentation in [63]. Due to limited data sample in addition to augmentation, k-fold cross-validation method was utilized in (n=8, 7.4%) studies [54], [55], [57], [64], [66], [69], [74], [83]. These cross-validation methods are also used to resolve issues such as overfitting and bias in dataset selection. Only one software named ITK-SNAP [159] was reported in [63] to segment structures in 3D images.

5.1.2 Fetus Head
This section contains three subsections (skull localization, brain standard planes, and brain disease). Total studies (n=43, 40.18%), but we only found one online public dataset called HC18 grand challenge [160]. This dataset was used in (n=13, 12.14%) studies [87]–[89], [91], [92], [95]–[98], [102], [104], [105], [161]. This dataset contains 2D US images collected from 551 pregnant women at different pregnancy trimesters. There were 999 images in the training set and 335 for testing; the sonographer manually annotated the HC. Image augmentation was applied in (n=9, 8.41%) studies [87], [89], [92], [96], [98], [102], [104], [105], [161], and cross-validation was employed in [161]. Table 4 highlights the best result achieved despite the challenges of recording the outcomes of the selected studies.
To detect and recognize the fetal NT based on 2D ultrasound images by
To support first trimester fetal assessment of multiple fetal anatomies
To localize the fetus and extract the best fetal biometry planes for the
to diagnose the placenta, amniotic fluid, and fetus.

To segment organs head, femur, and humerus in ultrasound images using multilayer super pixel images features
To classify abdominal fetal ultrasound video frames into standard AC
to segment the placenta, amniotic fluid, and fetus.

To evaluate the maturity of current Deep Learning classification techniques for their application in a real maternal-fetal medical environment.

To use the learnt visual attention maps to guide standard plane detection on all three standard anatomy planes: ADP, HCP, and FLP.

To support first trimester fetal assessment of multiple fetal anatomies including both visualization and the measurement of different structures.

To automatically classify different fetal structures in Z-D fetal ultrasound images by using information from both cropped regions of different fetal structures and the whole image.

To classify abdominal fetal ultrasound video frames into standard AC planes or background.

To detect of multiple fetal structures in free-hand ultrasound images.

To extract features from regions inside the images where meaningful structures exist.

To detect standard planes from US videos

Table 5. Articles published using AI to improve fetal body monitoring: Objective, backbone methods, optimization, fetal age best result, and observations.

| Study | Objective | Backbone Methods/Framework | Optimization/Extraction methods | Fetal age | Best Obtained Result | observations |
|-------|-----------|---------------------------|---------------------------------|----------|----------------------|--------------|
| [53]  | To identify the fetal skull, heart and abdomen from ultrasound images | SVM as the classifier | Gaussian Mixture Model (GMM) Fisher Vector (FV) | 20+ weeks | Acc: 0.9890 mAP: 0.9905 | 1) Methodology and result are clear. 2) Dataset acquisition is clear. 3) No data augmentation and no cross-validation. 4) Important performance metrics are missing. |
| [62]  | To segment the seven key structures of the neonatal hip joint | Neuronal Hip Rose Segmentation Network (NHSNet) | Feature Extraction Module Enhanced Dual Attention Module (EDAM) Two-Class Feature Fusion Module (2-Class FFM) Coordinate Convolution Output Head (CCOH) | 16 to 25 weeks. | DSC: 0.8705 HD: 0.82 mm AHD: 0.83 mm | 1) Multi-mechanisms were used to build the framework; the model is clear. 2) Data acquisition is precise, but data splitting is not cross-validated. 3) No data augmentation and no cross-validation method. 4) Novel work and unique in the target area. |
| [64]  | To segment organs head, femur, and humerus in ultrasound images using multilayer super pixel features images | Simple Linear Iterative Clustering (SLIC) Random forest | Unary pixel shape feature image moment | N/A | Femur Humerus Head: 0.90/0.98/0.91 F1: 0.531 0.458/0.357 Rec: 0.803 0.758/0.760 Spec: 0.972 0.988/0.9831 | 1) Methodology and results are not presented clearly. 2) Dataset is small; acquisition and splitting are not clear. 3) No data augmentation. 4) Cross-validation was applied. 5) First study that target fetus kidney segmentation. |
| [61]  | To evaluate the maturity of current Deep Learning classification techniques for their application in a real maternal-fetal medical environment | CNN DenseNet-169 | N/A | 18 to 40 weeks | Acc: 0.956 Error rate: 0.27 | 1. Methodology and results are not clear. 2. Dataset is large, data acquisition and splitting are precise. 3. No cross-validation was applied. 4. No comparison with state-of-the-art methods 5. Many misclassification performance metrics 6. Models have similar performance compared to an expert when classifying standard planes. |
| [57]  | To use the learnt visual attention maps to guide standard plane detection on all three standard anatomy planes: ADP, HCP, and FLP. | Temporal InceptionNet (TINet) Temporal attention module Convolutional LSTM Video classification module: Recurrent Neural Network (RNNs) | N/A | 20 to 40 weeks | DSC: 0.98 Jaccard index (J): 0.89 HD: 0.86 mm Mean Surface Distance (MSD): 0.24 mm | 1) Methodology and results are not presented clearly. 2) Dataset is small, acquisition and splitting are not clear. 3) No cross-validation was applied. 4) Novel work and unique in the target area. |
| [59]  | To automatically classify different fetal structures in Z-D fetal ultrasound images by using information from both cropped regions of different fetal structures and the whole image | FCNN: U-Net & UNET++ | N/A | 11 to 14 weeks | Acc: 0.915 Intersection over Union (IoU): 0.76 | 1) Clinical 3-D dataset was too small for a 3-D-FCN implementation even with data augmentation. 2) Performance metrics were not used such as F1 score and recall. |
| [58]  | To classify abdominal fetal ultrasound video frames into standard AC planes and background. | T-BNN framework LSTM | Features extracted using CNN classifier | 18 to 40 weeks | Mean Acc for all classes: 0.97 Pre: 0.76 | 1) Large dataset but no cross validation applied. 2) Paper is not organized in solid way 3) Role of SVM and Decision fusion were confusing for the reader. |
| [56]  | To classify abdominal fetal ultrasound video frames into standard AC planes and background. | M-SEB architecture Discriminator CNN | Generator CNN | N/A | Pre: 0.974 Rec: 0.962 F1: 0.961 | 1) Dataset acquisition is missing, the dataset small, and no cross-validation was applied. 2) Methodology is not clear 3) Novel and effective algorithm that simulates visual attention. |
| [54]  | To detect of multiple fetal structures in free-hand ultrasound images | Attention Gated LSTM | Class Activation Mapping (CAM) | 20 to 40 weeks | mean IOU is above 0.50 Average Pre: 0.94 Abdomen Average Pre: 0.93 Heart Average Prec: 0.85 | 1) Dataset acquisition and splitting is not apparent. 2) Only precision and IOU as performance metrics. |
| [55]  | To extract features from regions inside the images where meaningful structures exist. | Guided Random Forests | Probabilistic Boosting Tree (PBT) | 10 to 22 weeks | Mean Acc for all classes: 0.91 | 1) Dataset acquisition and splitting is not apparent 2) Only accuracy as performance metrics. |
| [60]  | To detect standard planes from US videos | T-BNN LSTMP (Transferred BNN) | Spatio-Temporal Feature J-CNN | 18 to 40 weeks | DSC: 0.80 PPV: 0.86 NPV: 0.86 Sens: 0.74 Spec: 0.77 F1: 0.74 Acc: 0.65 | 1) Largest dataset no cross-validation applied. 2) Incorrect comparison with other works due to different dataset. 3) T-BNN framework could be explained in a better way, such as form point 4. Same work in [59]. |
| [70]  | To propose the first and fully automatic framework in the field to simultaneously segment fetus, gestational sac and placenta. | 3D FCN + RNN hierarchical deep supervision mechanisms (HSDS) BiLSTM module denoted as FB-80 BiLSTM | Coefficient Conformity: 0.749 Average Distance of Boundaries: 0.925 RD of Blur: 9.978 | 10 to 14 weeks | DSC: 0.900 HD: 0.914 MM: 0.925 | 1) Methodology and results are very detailed but could be interpreted in better way. 2) Dataset is very small even after augmentation, data acquisition, and splitting are precise. |
| [71]  | To segment the placenta, amniotic fluid, and fetus. | FCNN | N/A | 11 to 19 weeks | DSC: 0.82 0.93 0.80 HD: 0.82 0.86 0.88 AD: 0.85 0.75 0.01 | 1) Methodology and results are very detailed, but multi-model experiments confuse the reader. 2) Dataset acquisition is clear, but data splitting is not precise. 3) No cross-validation was applied, no augmentation. |
| [72]  | To segment the amniotic fluid and fetal tissues in fetal US images | The encoder-decoder network based on VGG16 | CNN Structured Random Forests | N/A | 20+ weeks | Mean for all categories: Acc: 0.8099 Jaccard Index: 0.7803 | 1) Methodology is not explained well. 2) Dataset is very large after augmentation; data acquisition clear but splitting is not cross-validated. 3) Cross-validation was applied. 4) Different methods were compared. |
| [66]  | To detect fetal postures and the best post mortem planes for the head and abdomen from first trimester 3D fetal US images | CNN | Structural Random Forests | N/A | DSC: 0.901 HD: 0.901 MM: 0.901 | 1) Dataset acquisition is clear but no splitting is not precise. 2) Methodology steps are confusing for the reader. |
| [67]  | To segment the placenta, amniotic fluid, and fetus. | ResNet18 Soft Proposal Layer (SP) | ACC: 0.912 Jacc: 0.913 | 22 to 32 weeks | DSC: 0.90 | 1) Dataset acquisition is well explained but source and acquisition is not clear. 2) Validation is applied. 3) No cross-validation was applied. |
| [68]  | To reliably estimate abdominal circumference | CNN + Gradient Boosting Machine (GBM) Histogram of Oriented Gradient (HOG) | ACC: 0.933 (ma/60) Spec: 0.966 | 15 to 40 weeks | DSC: 0.90 | 1) Dataset acquisition is clear but not cross-validated. 2) Methodology is not organized well. 3) Validation is not applied. 4) Only Dice as performance metrics. |
| [69]  | To detect and recognize the fetal NT based on 2D ultrasound images by using artificial neural network techniques. | Artificial Neural Network (ANN) MultiLayer Perceptron (MLP) Network Bi-linear Interpolation Forward Propagation Method (BIFP) | N/A | ACC: 0.933 (ma/60) Spec: 0.966 | 4 to 12 weeks | Jaccard Index: 0.71 | 1) Dataset acquisition is not apparent. 2) Methodology very clear and explained in a consistent way. 3) Only Jacc and IoU were evaluated. |

(Continued)
### Growth disease

| Study | Objective | Backbone Methods/Framework | Optimization/Extractor methods | Fetal age | Best Obtained Result | observations |
|-------|-----------|----------------------------|-------------------------------|-----------|----------------------|--------------|
| [74]  | To propose the biometric measurement and classification of IUGR using OpenGL concepts for extracting the feature values and ANN model is designed for diagnosis and classification | ANN + Multi-layered perception neural networks (MLPNN) | Various | Various | ACC: 0.95 Root Mean Square Error (RMSE): 0.67 | 1) Methodology is not explained well. 2) Dataset is small, data acquisition and splitting is not precise. 3) Cross-validation was applied. 4) Two models were compared. |
| [75]  | To find the region of interest (ROI) of the fetal brain and diagnostic approach for monitoring fetal growth disease | CNN | N/A | 16 to 27 weeks | ACC: 0.9043 | 1) Dataset acquisition and splitting is not apparent. 2) Validation is not applied. 3) Only accuracy as performance metrics |
| [76]  | To detect fetal anomalies in 2-D US images | CNN | Various | 14 to 60 weeks | ACC: 0.60 | 1) Dataset acquisition and splitting is not apparent. 2) Validation is not applied. 3) Only Mean square error as performance metrics |
| [77]  | To develop a computer aided diagnosis and classification tool for detecting and PSO for Down Syndrome diagnosis. | CNN | Various | 14 to 40 weeks | ACC: 0.8162 | 1) Small dataset and dataset acquisition is not apparent. 2) Methodology could be explained in a better way. 3) Only accuracy as performance metrics. 4) The idea is novel but obtained very high accuracy on small dataset is suspicious. |

#### Gestational age (GA)

| Study | Objective | Backbone Methods/Framework | Optimization/Extractor methods | Fetal age | Best Obtained Result | observations |
|-------|-----------|----------------------------|-------------------------------|-----------|----------------------|--------------|
| [80]  | To automatically detect and measure the trans-cerebellar diameter (TCD) in the fetal brain, which enables the estimation of fetal gestational age (GA) | CNN FCN | Various | 16- to 26 weeks | ACC: 0.9789 loss: 0.0162 | 1) Methodology and results are not clear. 2) Dataset is sufficient after augmentation, data acquisition and is precise. 3) No cross-validation was applied. 4) No comparison with state-of-the-art methods. 5) Missing significance performance metrics. |
| [81]  | To accurately estimate the gestational age from the fetal lung region of US images | U-NET | Various | 24 to 60 weeks | Correlation coefficient: 0.76 P-value: 0.00045 | 1) Dataset acquisition is detailed but the dataset small, and no cross-validation is applied 2) Missing essential performance metrics. 3) Methodology was not clear for the reader. |
| [82]  | To classify, segment, and measure several fetal structures for the purpose of GA estimation | U-NET | ResNet | 16+ weeks | ACC: 0.91 loss: 0.91 Mean Absolute Error: 1.89 centimeters | 1) Dataset acquisition and splitting is not apparent, and no cross-validation is applied 2) Methodology steps are clear and organized |

#### Gender identification

| Study | Objective | Backbone Methods/Framework | Optimization/Extractor methods | Fetal age | Best Obtained Result | observations |
|-------|-----------|----------------------------|-------------------------------|-----------|----------------------|--------------|
| [83]  | To measure the accuracy of Learning Vector Quantization (LVQ) to classify the gender of the fetus in the US image | ANN | Various | Various | ACC: 0.8089 Mean Absolute Error (MAE): 0.3089. Rec: 0.611 Pre: 0.55 F1: 0.556* | 1) Methodology is not clear, but obtained result is clearly explained. 2) Dataset is very small, data acquisition and splitting are not precise. 3) Cross-validation was applied, no augmentation. 4) Different models were compared. 5) Novel work by identifying gender in early stage of pregnancy. |
Skull localization and measurement

[07] To locate the fetal head region in US imaging
Multi-scale mini-LinkNet network
N/A
12 and 40 weeks
DSC: 0.995
1) methodology is clear. 2) result is clear for reader. 3) no cross validation is applied

[08] To locate the fetal head from 3D ultrasound images using shape model
Adultrost
Shape Model
Marginal Space
11 to 14 weeks
ACC: 0.83
1) methodology could be interpreted in better way. 2) dataset very small. 3) only accuracy as performance metrics.

[09] To detect fetal head
Deep Bilinear Network (DBN)
Restricted Boltzmann Machines
Hough transform
Histogram Equalization
11 to 14 weeks
ACC: 0.86
1) Methodology is clear. 2) Only clinical dataset was reported and synthetic data were ignored. 3) important performance metrics are missing.

[10] To automatically segment fetal head from maternal and other fetal tissue
U-Net
Eclipse timing
12 and 20 weeks
ACC: 0.83
1) methodology is clear. 2) dataset very small even public dataset is available. 3) important performance metrics are missing such as DSC

[103] To automatically discover and localize anatomical landmarks, measure the HC, TV, and the TC
CNN
Salinity maps
13 to 26 weeks
CUP/LV/Carot HC TV alignment errors: 9.8 4.1 7.1
1) Methodology and results are confusing for the reader.
2) Dataset acquisition and splitting are not clearly defined. 3) Important performance metrics are missing.

[105] To demonstrate the effectiveness of hybrid method to segment fetal head
BD-Net
Scattering Coefficients (SC)
13 to 26 weeks
DSC: 0.733
HD: 1.58 mm
1) methodology could be explained better for the reader. 2) dataset splitting is not apparent, and no validation. 3) proposed method did not achieve the best result.

[106] To segment fetal head using Network Binarization
DeepSeperable Convolutional Neural Networks (DSCNN)
Network Binarization
12 and 40 weeks
DSC: 0.8689
1) Methodology orientation is not structured. 2) no validation. 3) important performance metrics are missing.

[107] To segment the fetal skull boundary and fetal skull for HC measurement
U-Net
Square and Extinction (SE) blocks
12 and 40 weeks
DSC: 0.7931
1) methodology orientation is clear. 2) slightly improvement compared with original model. 3) other performance metrics are missing.

[108] To automatically segment and estimate of HC, ellipsoid.
Multi-Yolk network based on Link-Net architecture (MTLA)
Ellipse Funnor
12 and 40 weeks
DSC: 0.5684
1) Methodology and result are clear. 2) Proposal method did not obtain the best result compared to the current work.

[109] To capture more information with multiple-channel convolution from US images
Multiple-Channel and Atrous Mix-Net
Encoder and Decoder Module
N/A
Dice: 0.971
Rec: 0.978
DSC: 0.812
Mean Pixel ACC (mPA): 98.02
1) Methodology and result are clear and detailed. 2) dataset acquisition is clearly explained no validation is applied. 3) proposal method achieved the best result compared to other works except in precision.

[109] To automatic segment fetal ultrasound image and HCometry
Deeply Supervised Attention-Gated (DAG)-V-Net
Attention-Gated Module
12 and 40 weeks
DSC: 0.9793
AUPR: 0.877
HD: 0.69 mm
HD: 1.27 mm
1) methodology and result are clear and detailed. 2) dataset acquisition is clearly explained compared to other state-of-the-art methods except in DSC. 3) provided many potential applications to deal with the actual Ultrasound dataset.

[110] To compose a new US volume containing the whole brain anatomy
U-Net + Incidence Angle Maps (IAMS)
CNN
Normalized Mutual Information (NMI)
13 to 26 weeks
DSC: 0.871
1) methodology and result are confusing for the reader. 2) dataset acquisition and splitting are not clearly defined. 3) paper organization is poor. 4) comparison with state-of-the-art methods.

[111] To directly measure the head circumference, without having to resort the handcrafted features or manually labeled segmented images
CNN regressor (Reg-ResNet50)
N/A
12 to 40 weeks
MSE: 36.21 pixels MRE: 34.78 pixels
1) Methodology and result are simply explained and understandable. 2) Dataset acquisition and splitting are clearly defined. 3) Paper organization is poor. 4) First attempt to use CNN as a regressor to deal with fetal HC without segmentation.

[115] To propose regime-CNN for head localisation and centering, and a regression CNN for accurately delineate the HC
CNN regressor (U-net)
DeepVNet2
12 to 40 weeks
DSC: 0.9776
AUPR: 1.00
DF: 0.21 mm
HD: 0.32 mm
1) methodology and result are simply explained and understandable. 2) dataset acquisition and splitting are clearly defined. 3) Second attempt to use CNN as a regressor to deal with fetal HC without segmentation.

[112] To present a novel end-to-end deep learning network to automatically measure the fetal HC, biplanar diameter (BPD), and occipito-mental diameter (OMD) length from 2D US images
FSNet (SAPNet)
Regression network
12 to 40 weeks
DSC: 0.972
AUPR: 0.03 mm
Mean Pixel ACC (mPA)
1) methodology is clear, result has multiple comparisons confused the reader. 2) Dataset volume after segmentation were not reported. 3) regression network was not clarified.

[114] To segment fetal head from US images
FCN
12 to 40 weeks
DSC: 0.9773
AUPR: 0.877
DF: 1.49 mm
HD: 1.39 mm
1) methodology is precise. 2) dataset after augmentation is not reported. 3) No validation method, no comparison.

[118] To deal with a completely computerized detection device of next fetal head
Multi-Yolk network based on Link-Net architecture (MTLA)
Hedged-Transform (HT)
ANN Feed Forward (NFFV) Classifier
N/A
ACC: 0.79
1) methodology and result is not precise. 2) Dataset acquisition and splitting are not clearly defined. 3) paper is not organized

[119] To measure HC, automatically
Random Forest Classifier
Head-like features
EllisFit method
18 to 33 weeks
DSC: 0.966
Prec: 0.968
Rec: 0.967
F1: 0.960
1) methodology and result are precise. 2) dataset is small with no augmentation and data acquisition is not detailed. 3) no validation method. 4) only accuracy is reported 5) paper is not organized

[120] To determine measurements of fetal HC and BPD
RCR
N/A
18 to 22 weeks
ACC: 0.85
1) methodology is clear. 2) Dataset in large with no augmentation, and data acquisition is detailed 3) comparison between expert annotation and model prediction

[121] To segment the whole fetal head in US volumes
Hybrid attention scheme (HASS)
3D-UNet - Encoder and Decoder architecture for dense labeling
20 to 31 weeks
DSC: 0.9605
Prec: 0.942
BD boundaries: 4.60 mm
1) methodology and result are precise. 2) dataset is small but augmented to obtain high volume. 3) comparison is provided with another model. 4) claimed to be the first investigation about whole fetal head segmentation in 3D US, but many works additionally proposed 5) same dataset was used. 6) comparison was not fair. 7) paper was overwhelmed with unnecessary details.

[123] To segment the fetal head using a flexibly plug-in-play module called vector self-attention layer (V-SAL)
CNN
Vector Self-Attention Layer (VSAL)
Context Aggregation Loss (CAL)
12 and 40 weeks
DSC: 0.791
PB ACC (PA): 0.69
1) methodology and result are precise and detailed. 2) dataset in public, no augmentation is used. 3) Comparison is provided with another models. 4) Paper was overwhelmed with unnecessary details

[124] To provide automatic framework for skull segmentation in fetal 3D US
Two-stage Cascade CNN (2S-CCNN) U-Net
Incidence Angle Map Shadow Casting Map
20 to 36 weeks
DSC: 0.83
Symmetric Surface Distance (SSD): 0.19 mm
1) methodology and result are precise and detailed. 2) dataset in public, no augmentation is used. 3) Comparison is provided with another models. 4) Paper was overwhelmed with unnecessary details

[125] To segment 2D ultrasound images of fetal heads based on a V-Net
Fully Convolutional Neural Network Combination vector
N/A
12 to 40 weeks
DSC: 0.797
Jaccard index [0.5 0.70]
[0.5 0.95]
ACC: 0.988
1) methodology and result are precise. 2) Dataset is very small, the volume of data was not reported after the segmentation. 3) no validation method. 4) claimed to be the first investigation about whole fetal head segmentation in 3D US, but many works additionally proposed 5) same dataset was used. 6) comparison was not fair. 7) Paper could be organized in better way.

[126] To segment the cranial pixel in an ultrasound image using a random forest classifier
Random Forest Classifier
Simple Linear Iterative Clustering (SLIC) Head Features
25 to 34 weeks
Prec: 0.999
Rec: 98.11
ACC: 0.9722
1) methodology and result are precise. 2) dataset is very small with no augmentation, and data acquisition and splitting are not clear. 3) No comparison with other methods. 4) Paper is poorly written and incoherent

[127] To automatically estimate fetal HC
U-Net
Monte-Carlo Dropout
18 to 22 weeks
DSC: 0.802
AUPR: 0.81
1) methodology and result are not precise. 2) Dataset acquisition and splitting is not clear. 3) no validation method.

Brain standard plane

[107] To automatically recognize six standard planes of fetal brains.
CNN + Transfer learning ECNN
N/A
18 to 22 weeks
ACC: 0.971
Pre: 0.955
F1: 0.900
1) clear methodology. 2) Multi dataset are used with data augmentation

[114] To help the clinician or sonographer obtaining those planes of interest by finding the fetal head alignment in 3D US
Random forest classifier
Shape model and template deformation algorithm Hough transforms
19 to 28 weeks
5.1
1) Methodology is not clear many objective in one flow
2) dataset very small
3) Only error rate as performance metrics.
To develop computer-aided diagnosis algorithms for five common fetal brain abnormalities:

| Study | Objective | Backbone Methods/Framework | Optimization/Extractor methods | Fetal age | Best Obtained Result | observations |
|-------|-----------|---------------------------|--------------------------------|-----------|----------------------|-------------|
| [109] | To segment the lateral ventricle from 2D US images. | U-Net+ResNet (U-Net-C) | N/A | 18 to 20 weeks | Dice: 0.97 | 1) Methodology and results are presented clearly. 2) Dataset is small but acquisition and splitting are clear. 3) Important performance metrics are used |
| [116] | To detect multiple planes simultaneously in challenging 3D US datasets | Multi-Task Retriment Learning (MARTL) | RNN | 19 to 31 weeks | Bland Altman discrepancy between the difference between the ED (mm) and the LV volume origin (mm) (Mean: 1.19, SD: 0.86) | 1) Methodology and result detailed written 2) Second work that used of AI, with the fetal dataset 3) Some details are not necessary to be included 4) Such as great work need future work 5) Some performance metrics are missing |
| [119] | To detect standard plane and quality assessment | Multi-task learning Framework Faster R-CNN (MH-F-R-CNN) | N/A | 14 to 20 weeks | DSC: 0.9494, ACC: 0.9255, Spec: 0.9106 | 1) Methodology and result are precise and detailed. 2) Dataset is transparent in terms of acquisition and splitting. 3) Classification module should adopt only one architecture instead of seven. 4) Detection only evaluated using mean average precision, where mean average precision were used for classification |
| [119] | To tackle the automated problem of fetal biom | U-Net, CNN | N/A | 18 to 20 weeks | Dice: 0.9143, ACC: 0.9539 | 1) Methodology and result are not precise 2) Dataset is very small but transparent in terms of acquisition and splitting 3) Only accuracy and DSC were used as performance evaluation |
| [117] | To determine the standard plane in US images | Faster R-CNN | N/A | 14 to 20 weeks | DSC: 0.9493, ACC: 0.9720 | 1) Methodology and result are precise. 2) Dataset acquisition is not detailed 3) No validation method. 4) Only Average Precision (AP) is reported |
| [112] | To address the problem of 3D fetal brain localization, structural segmentation, and alignment to a referential coordinate system | Multi-Task PCN | Slice-Wise Classification | 10 to 34 weeks | Average Center Deviation: 2.0 mm Average Scale Deviation: 2.3 | 1) Methodology and results are precise and detailed 2) Dataset is small with no augmentation 3) No comparison has been provided for work in great work 4) Performance evaluation metrics should be used such as AUC, DSC, etc. |}

### Brain Differences

- **Methodology and results are clear.**
- **Dataset acquisition and splitting are clear.**
- **Important performance metrics are used.**
- **No validation method.**
- **First work that use RL to localize fetal head.**

### Comments

- **1) Methodology and result are not precise.**
- **Dataset is very small and transparent in terms of acquisition and splitting.**
- **No validation method.**
- **First work that use RL to localize fetal head.**
Face anatomical landmarks

To detect position and orientation of facial region and landmarks

- SFPD-Net (Samsung Fetal Face Detection Network)
- CNN Backdoor Network
- Region Proposal Network (RPN)

To detect nasal bone for US of fetus

- Local Binary Patterns (LBP)
- Histogram of Oriented Gradient (HOG)

Face expressions

To recognize facial expressions from 3D US

- ANN
- Histogram equalization
- Thresholding
- Morphing
- Sampling
- Clustering
- Local Binary Patterns (LBP)
- Minimum Bounding and Maximum Relevance (MBMR)

To recognize fetal facial expressions that are considered as being related to the brain development of fetuses

- CNN
- Histogram of Oriented Gradient (HOG)

Table 7. Articles published using AI to improve fetus face monitoring: Objective, backbone methods, optimization, fetal age best result, and observations

| Study | Objective | Backbone Methods/Framework | Optimization/ Extractor methods | Fetal age | Best Obtained Result | Observations |
|-------|-----------|----------------------------|---------------------------------|-----------|----------------------|-------------|
| [135] | To address the issue of recognition of standard planes (i.e., nasal, coronal and sagittal planes) in the fetal ultrasound (US) image | SVM classifier | AdultNet for detect region of interest, RID | 20 to 36 weeks | Mean Average Precision (mAP): 0.9995 | Method and result are not clear. 2) Dataset is small, and the acquisition and splitting was not clear. 3) No data augmentation was not reported. 4) Other evaluation methods were not mentioned clearly. |
| [126] | To automatically recognize the FSPF from US images | Deep convolutional networks (DCNs) | N/A | 20 to 36 weeks | Acc: 0.9999  Rec: 0.9999 | Method and result are clear. 2) Dataset is small, and acquisition and splitting was not clear. 3) No data augmentation was not reported. 4) No validation method |
| [127] | To automatically recognize FSPF via a deep convolutional neural network (DCNN) architecture | DCNN | t-Distributed Stochastic Neighbor Embedding (t-SNE) | 20 to 36 weeks | Acc: 0.9959  Rec: 0.97 | Method and result are clear. 2) Dataset is small, and acquisition and splitting was not clear. 3) Data augmentation and validation was employed. 4) DCNN outperform other methods. |
| [128] | To automatic recognition of the fetal facial standard planes (FSPS) | SVM classifier | Root scale invariant feature transform (RootSIFT) Gaussian mixture model (GMM) Fisher Vector (FV) Principle Component Analysis (PCA) | 20 to 36 weeks | Acc: 0.9327  mAP: 0.9919 | Multi-techniques were used in methodology and result is clear, detailed, and consistent. 2) Dataset acquisition and splitting was not clear. 3) Data augmentation and validation is not employed. 4) No comparison with other works. 5) For such great and detailed work, only precision and accuracy were reported. |
| [129] | To automatic recognize and classification of FSPS | SVM classifier | Local Binary Patterns (LBP)  Histogram of Oriented Gradient (HOG) | 20 to 24 weeks | Acc: 0.9467  Rec: 0.9388  Pre: 0.9467  F1: 0.9480 | Methodology and result are consistent. 2) Dataset splitting was not clear. 3) No data augmentation. 4) No comparison with other works. |

Table 8. Articles published using AI to improve fetus heart monitoring: Objective, backbone methods, optimization, fetal age best result, and observations

| Study | Objective | Backbone Methods/Framework | Optimization/ Extractor methods | Fetal age | Best Obtained Result | Observations |
|-------|-----------|----------------------------|---------------------------------|-----------|----------------------|-------------|
| [130] | To detect position and orientation of facial region and landmarks | SFPD-Net (Samsung Fetal Face Detection Network) multi-class segmented | N/A | 14 to 30 weeks | L-Eye: L2 norm (mm): 2.21  R-Eye: L2 norm: 1.56  Nose L2 norm: 1.31 | Method and result are clear. 2) Dataset is small, and the acquisition and splitting was not clear. 3) Volume after data augmentation was not reported. 4) Other evaluation methods were not mentioned. |
| [131] | To detect landmarks in 3D fetal US volumes | CNN Backdoor Network | Region Proposal Network (RPN) Bounding-box regression | N/A | AP: 0.250  mIoU: 0.6377 | Method and result are clear. 2) Dataset is small, and acquisition and splitting was not clear. 3) Volume after data augmentation was not reported. 4) No validation method. 5) Other evaluation methods were not mentioned. |
| [132] | To detect nasal bone for US of fetus | Back Propagation Neural Network (RBFNN) | Bone Cone Transform (BCT)  Bauchhoeche 04 Wavelet transform | 11 to 13 weeks | Acc: 0.906  mIoU: 0.6377 | Methodology and result are inconsistent. 2) Dataset acquisition and splitting was not clear. 3) No data augmentation and validation are not employed. 4) No comparison with other works. |
| [133] | To detect position and orientation of facial region and landmarks | CNN | t-Distributed Stochastic Neighbor Embedding (t-SNE) | N/A | Acc: 0.971  Rec: 0.85  Pre: 0.88 | Methodology, multi techniques were used. However, they can be replaced by using CNN. 2) Dataset is small, and the acquisition and splitting was not clear. 3) Data augmentation was not reported. 4) No validation method. |
| [134] | To recognize facial expressions from 3D US | ANN | Histogram equalization  Thresholding  Morphing  Sampling  Clustering  Local Binary Patterns (LBP)  Minimum Bounding and Maximum Relevance (MBMR) | N/A | Acc: 0.91  Rec: 0.85  Pre: 0.84 | Methodology and result were very detailed but confusing for the reader and could be written and organized consistent way. 2) Dataset is sufficient after the use of data augmentation. 3) Only ACC and AUC are reported. 4) Novel work and significant for clinical use. |
| [135] | To recognize facial expressions that are considered as being related to the brain development of fetuses | CNN | Histogram of Oriented Gradient (HOG) | N/A | Acc: 0.965  Rec: 0.996  Spec: 0.993 | Methodology and result are clear. 2) Many techniques were used to build the framework. 3) Dataset acquisition splitting is clear. 4) No data augmentation. |
| [136] | To detect cardiac substructures and structural abnormalities in fetal US videos | U-Net + DeepLabV3+ | N/A | N/A | Dice: 0.997  Cross Entropy Loss: 0.018  mIoU: 0.720 | Methodology and results are very detailed but confusing for the reader and could be written and organized consistent way. 2) Dataset is sufficient after the use of data augmentation. 3) Only ACC and AUC are reported. 4) Novel work and significant for clinical use. |
| [137] | To perform multi-disease augmentation and multi-class semantic segmentation for the five key components | U-Net + DeepLabV3+ | N/A | N/A | Dice: 0.997  Cross Entropy Loss: 0.018  mIoU: 0.720 | Methodology and results are very detailed but confusing for the reader and could be written and organized consistent way. 2) Dataset is sufficient after the use of data augmentation. 3) Only ACC and AUC are reported. 4) Novel work and significant for clinical use. |
| [138] | To recognize and judge fetal congenital heart disease (FHD) development | DGACNN Framework | CNN  Wasserstein-GAN + Gradient Penalty  90GAN-GP  DANomaly Faster-RCNN | 18-39 weeks | Acc: 0.856  mAP: 0.851 | Methodology and results are very detailed but confusing for the reader and could be written and organized consistent way. 2) Dataset is sufficient after the use of data augmentation. 3) Only ACC and AUC are reported. 4) Novel work and significant for clinical use. |
| [139] | To segment the ventricular septum in US | Crupping-Segmentation-Calibration (CSC) | Volumetry cropping module  U-Net augmentation  Male  VGG-backbone Calibration Module | 18 to 28 weeks | mIoU: 0.897 | Methodology and results are clear. 2) Many techniques were used to build the framework. 3) Dataset acquisition splitting is clear. 4) No data augmentation. |
| [140] | To detect cardiac substructures and structural abnormalities in fetal US videos | Supervised Object detection with Normal Data Only (SODNO) | CNN  VGG0v2 | 18 to 34 weeks | Heart mean AUC: 0.976  Vessels: 0.991 | Two techniques were integrated to build the framework. 2) Methodology and results are clear. 3) Large dataset and no augmentation needed. 4) Supplement materials were added, including real-time record. 5) Significant work and first application were seen in this survey. |
| [131] | To identify recommended cardiac views and distinguish between normal hearts and complex CHD and calculate standard fetal cardiothoracic measurements | Ensemble of Neural Networks | ResNet and U-Net Seg-3D-GAN | 18 to 24 weeks | Acc: 0.970  Spec: 0.95 | Methodology and results are very detailed, but a paper’s organization is difficult for the reader. 3) Largest dataset in the field, acquisition, and splitting is clear 4) No comparison with other works. 5) Fromaging result was achieved. |
| [143] | To learn the features of Holosylcin Intracardiac Focus (HIF) that can cause Down Syndrome (DS) whereas testing phase classifies the EFL into DS positive or DS negative based | Multi-scale Quantized Convolution Neural Network (MQCNN) | Cross-Correlation Technique (CCT)  Enhanced Learning Vector Quantizer (ESLVQ) | 24-26 weeks | Acc: 0.99  Spec: 0.99  Rec: 0.99 | Methodology and result are not explained clearly. 2) Very small dataset, acquisition, and splitting was not clear. 3) No validation method, no data augmentation. 4) Evaluation matrices were not accurately reported. |
| [144] | To perform automated diagnosis of hypoplastic left heart syndrome (HLHS) | SimNet (VGG16) | N/A | 18-22 weeks | Acc: 0.93  Rec: 0.90  Pre: 0.83  F1: 0.87 | Methodology is based on other work [144] result is explained clearly. 2) Large dataset but data acquisition, and splitting was not clear 3) No validation method, no data augmentation. 4) No comparison with other works. |
Table 8. Continued

| Study | Objective | Backbone Methods/Framework | Optimization/Extractor methods | Fetal age | Best Obtained Result | observations |
|-------|-----------|---------------------------|--------------------------------|----------|----------------------|-------------|
| [143] | To perform automated segmentation of cardiac structures | U-Net | Structural Similarity Index Measure (SSIM) | N/A | DSC: 0.856 | 1) Methodology and results are clear. 2) No data augmentation and no validation method. 3) Dataset acquisition splitting is clear. |
| [144] | To accurately segment seven important anatomical structures in the A4C view | DW-Net | Dilated Convolutional-Chain (DCC) module | N/A | DSC: 0.827 | 1) Methodology and results are clear. 2) Two techniques were used to build the framework. 3) Dataset splitting is not clear. 4) No data augmentation. 5) Largest time that a CNN based deep learning method is applied in the segmentation of multiple anatomical structures in A4C views. |
| [148] | To automatically control the fetal US heart chamber plane | CGNN-based Framework | Basic CNN, a variant of DenselyNet Deep CNN with 3Dmax pool as basic architecture. The ARVNet for real-time object detection. | 14 to 20 weeks | ACC: 0.9849 | 1) Three different networks were used to build the framework. 2) Dataset splitting is not clear. 3) Each network was tested separately, and the average performance for the whole framework was not reported. |
| [147] | To localize the end-systolic (ES) and end-diastolic (ED) from ultrasound | Hybrid CNN based framework | TopoV3 Maximum Difference Filtering (MDF) Transferred CNN | 10 to 36 weeks | ACC: 0.993 | 1) Many different techniques were integrative to build the framework. 2) Volume of the dataset after augmentation was not reported. 3) No validation method. 4) Significant work and first-time MotifNet framework were used to deal with the fetal US and achieved promising results. |
| [145] | To detect the fetal heart and classifying each individual frame as belonging to one of the standard viewing planes | FCN | N/A | 20 to 35 weeks | Segmentation error rate: 0.2348 | 1) Methodology and result are clear. 2) Large dataset and volume after augmentation is not reported. 3) Only error rate is reported and no validation method. |
| [142] | To jointly predict the viability, view plane, location of the fetal heart in US videos. | Multi-Task CNN | Hierarchical Temporal Encoding (HTE) | 20 to 35 weeks | Classification: 0.6558 Localisation: 0.7968 | 1) Methodology is not clearly explained. 2) Dataset acquisition and splitting were not clear. 3) Only accuracy rate is reported and no validation method. |

Table 9. Articles published using AI to improve fetus abdomen monitoring: method details, objectives, health issue, dataset details and model observations

| Study | Objective | Backbone Methods/Framework | Optimization/Extractor methods | Fetal age | Best Obtained Result | observations |
|-------|-----------|---------------------------|--------------------------------|----------|----------------------|-------------|
| [148] | To automatically detect two anatomical landmarks in an abdominal image plane stomach bubble (SB) and the umbilical vein (UV). | Adaboost | H-3-like feature | 14 to 19 weeks | SB ACC: 0.70 | 1) Methodology is explained clearly but result need more details. 2) Dataset acquisition and splitting was clear. 3) No validation method, no data augmentation. 4) No comparison with other works. |
| [149] | To localize abdominal standard plane (ASPA) from US including SB, UV, and spine (SP) | Random Forest Classifier+ SVM | H-3-like feature | 16 to 40 weeks | Overall ACC: 0.816 | 1) Methodology and result are not clearly explained. 2) Dataset acquisition and splitting was not clear. 3) No validation method, no data augmentation. 4) Only detection rate was used to evaluate the performance. |
| [155] | To classify ultrasound images (SB, amniotic fluid (AF), and UV) and to obtain an initial estimate of the AC | Initial Estimation CNN+U-Net | Hough transform | N/A | ACC: 0.9255 Abdominal Plane acceptance ACC: 0.871 | 1) Many networks were used to build three integrated frameworks, methodology and results are confusing for the reader. 2) Dataset acquisition and splitting was not clear. 3) No validation method. 4) Evaluation metrics were not used clearly. |
| [157] | To classify ultrasound images (SB, AF, and UV) and measure AC | CNN | Hough transform | 20 to 34 weeks | ACC: 0.919 Abdominal Plane acceptance ACC: 0.809 | 1) Many networks were used to build the three integrated frameworks, methodology and results are confusing for the reader. 2) Dataset acquisition and splitting was not clear. 3) No validation method. 4) Evaluation metrics were not used clearly. |
| [150] | To find the region of interest (ROI) of the fetal abdominal region in the US image. | Fetal US Image Quality Assessment (FIQUA) | L-CNN is able to localize the fetal abdomen ROI. U-Net CNN then further analyzes the identified ROI. DCCN to duplicate the US images for the RGB channels. | 16 to 40 weeks | SB ACC: 0.70 UV ACC: 0.996 | 1) Many networks were used to build the FIQUA framework. 2) Methodology and results are confusing for the reader. 3) Dataset splitting was not clear. 4) No validation method. 5) No comparison with other works. |
| [151] | To localize the fetal abdominal standard plane from ultrasound | Random Forest Classifier+ SVM classifier | Radial Component-Based Model (RCBM) Vessel Probability Map (VPM) | 16 to 40 weeks | SB ACC: 0.999 SP ACC: 0.999 UV ACC: 0.996 | 1) Two classifiers were used to improve detection speed and accuracy. 2) Methodology is not consistently explained. 3) Dataset splitting was not clear after data augmentation. 4) No validation method. |
| [156] | To diagnose the (prenatal) US images by design and implement a novel framework | Defending Against Child Death (DADCI) | CNN U-Net | Hough-man transformation | N/A | ACC: 0.999 SB ACC: 0.999 UV ACC: 0.996 | 1) Classification integrated with segmentation was used to improve detection speed and accuracy. 2) Methodology is not consistently explained. 3) Dataset splitting was not precise, no augmentation. 4) No validation method. 5) Some inconsistency during the comparison with other works, however, U-Net provides plausible results. |
| [152] | To detect important landmarks employed in manual scoring of ultrasound images. | Adaboost | H-3-like feature | 16 to 27 weeks | ACC: 0.900 UV ACC: 0.996 | 1) Methodology and result is clear but not detailed 2) Dataset acquisition and splitting are missing; no augmentation. 3) No validation method. 4) No comparison with other works. |
| [153] | To automatically detect the end-systolic (ES) and end-diastolic (ED) from ultrasound images. | Adaboost | One Combined Vessel Classifier (DVCS) Two Separately Trained Classifiers (2S2C) H-3-like feature | 20 to 28 weeks | Rec: 0.7129 Pre: 0.7829 | 1) Methodology is clear but result missing more detailed 2) Dataset acquisition and splitting are missing. 3) No validation method. 4) No comparison with other works. 5) First work in the area. |
| [154] | To localize the FASP from US images. | DVCN | Fine-Tuning with Knowledge Transfer Barnes-Shoer-Stochastic Neighbor Embedding (BS-SNE) | 18 to 40 weeks | ACC: 0.910 | 1) Methodology is not clear. 2) Dataset acquisition and splitting are missing; data augmentation was reported. 3) No validation method. 4) Comparison with other works (5) only accuracy was reported. |


6. Discussion

6.1 Principal findings

AI techniques applied are shown in Figure 4. Deep learning (DL) was the most utilized techniques, seen in (n=81, 75.70%) studies. Traditional Machine learning (ML) models are used in (n=16, 14.95%) studies. Artificial neural network (ANN) models are only used in (n=7, 6.5%) studies. We also found the use of Reinforcement learning (RL) in (n=2, 1.86%) studies. Both deep learning as well as machine learning models were used in one study.

- Deep learning for fetus health

As seen in Figure 4, DL was used heavily in all fetal organ experiments including general fetus issues, fetus head, fetus face, fetus heart, and fetus abdomen. Furthermore, as seen in most of the models (n=74, 69.1%), Convolutional Neural Network (CNN) used as a backbone of deep learning and is used in all tasks including classification, segmentation and object detection [54], [56], [57], [59], [63], [65]–[68], [72], [73], [76], [77], [81], [82], [84], [87]–[91], [93]–[99], [102]–[107], [109], [110], [113], [117]–[124], [126], [127], [130], [131], [134]–[144], [146], [147], [150], [153], [155]–[157], [161], [163]. Additionally, Fully Connected Neural Networks (FCNNs) is similarly largely utilized as a backbone or part of the framework (n=11, 10.28%) of DL, as seen in [63], [65], [70], [71], [80], [92], [100], [104], [112], [145], [146]. In addition to CNN and FCN, a Recurrent Neural Network (RNN) was used to perform classification tasks in hybrid frameworks as seen in (n=5, 4.67%) studies [54], [57], [59], [60], [70]. U-Net is a convolutional neural network that was developed for biomedical image segmentation. Therefore, we found U-NET to be the most frequently utilized segmentation model as a baseline of the framework or as optimization. This was seen in (n=23, 21.50%) [63], [73], [81], [82], [88], [90], [94], [95], [97], [101], [109], [113], [118], [119], [122], [137], [138], [141], [143], [144], [155], [156], [163]. Other segmentation models were also used, such as DeepLabV3 in [137], LinkNet in [106], and the Encoder-Decoder network based on VGG16 in [72]. A Residual Neural Network (ResNet) was used to perform some tasks and enhance the framework efficiency as seen in (n=6, 5.60%) studies [62], [67], [82], [109], [110], [123], [138]. Some popular object detection models (n=12, 11.21%) were utilized to perform detection tasks in the fetus head, face, and heart. Additionally, Region Proposal Network (RPN) with bounding-box regression technique was used in [119], [131]. Faster R-CNN was utilized in [104], [117]–[119]. Further, YOLO (You Only Look Once) models were utilized; YOLOv2 in [105], [140], [141] and YOLOv3 in [147]. Mask R-CNN was utilized only in [124] and Aggregated Residual Visual Block (ARVB) was used only in [146]. Surprisingly, we have found CNN is used for regression to predict the accurate measurement of HC as seen in [102], [103], [105], [161]. We found that, due to the noise and characteristics of US images, DL still relies on traditional feature extraction techniques in image analysis used to optimize the DL frameworks for better performance. For example, Hough transform was utilized with head and abdomen US images as seen in [84], [155]–[157]. Additionally, image processing techniques such as histogram equalization was used in [84]. Besides these tools, Class Activation Maps (CAM) are a helpful classification tool in computer vision that was utilized for brain visualization in [122], [123], fetal part structure in [54], and heart chamber view in [138]. DL shows promising results for the identification of certain fetus diseases as seen in the identification of heart diseases [135]–[141], brain diseases [120]–[124], and growth diseases [75]–[77]. Table 10 list of DL studies exhibiting technical novelty that achieved promising results in each category.

| Main Organ               | Sub-section               | Best study                                      |
|--------------------------|---------------------------|------------------------------------------------|
| Fetal Body               | Fetal part structures     | [62], [65]                                     |
|                          | Anatomical structures     | [67], [73]                                     |
|                          | Growth disease            | [77]                                           |
|                          | Gestational age           | [81]                                           |
|                          | Skull localization        | [105], [161]                                  |
| Head                     | Brain Standard plan       | [107], [108], [110]                            |
|                          | Brain disease             | [120], [124]                                  |
|                          | Fetal facial standard     | [127]                                          |
| Face                     | Face anatomical           | [130]                                          |
|                          | Facial expression         | [134]                                          |
| Heart                    | Heart disease             | [135], [137], [138], [140]                     |
|                          | Hear chamber view         | [144], [147]                                  |
| Abdomen                  | Abdominal anatomical      | [155], [156]                                  |
Most of the research utilizing machine learning for fetal health monitoring was conducted between 2010 and 2015, as seen in (n=14, 13.8%). Only two studies [85], [129] were published in 2018 and 2021, respectively. We found that ML models were used at all four stages except for the heart and identify the phenotype of fetus disease. The most used ML algorithm was Random Forest (RF), as seen in (n=8, 7.47%) studies. RF was used as a baseline classifier in [85], [93], [111], [114]. RF was used with other classifiers as well, with Support Vector Machine (SVM) in [149], [151] and Probit Tree (PTB) in [55]. Simple Linear Iterative Clustering (SLIC) was also used in [64]. The SVM classifier was used in (n=6, 5.60%) studies. SVM was used as a baseline classifier in [53], [125], [128], [129] and utilized with other classifiers as seen in [149], [151]. AdaBoost was used in (n=5, 4.67%) studies as a baseline classifier in [86], [148], [152], [154] and to detect the region of interest (ROI) in [125]. Moreover, ML traditional algorithm still rely on Haar-like features (digital image features used in object recognition) while being utilized with US images as seen in (n=9, 8.41%) studies [65], [86], [93], [111], [148], [149], [151], [152], [154]. Gaussian Mixture Model (GMM) and Fisher Vector (FV) were used in (n=3, 2.80%) studies, where GMM is used to simulate the distribution of extracted characteristics throughout the images. Then, FV represents the gradients of the features under the GMM, with respect to the GMM parameters as seen in [53], [118], [125].

Dubbed ML were seen in [53] with the SVM with decision fusion for classification and features extracted by fine-tuning AlexNet. Finally, we found the first research to utilize ML model in monitoring fetus was in [154], which used to classify abdominal anatomical landmarks.

### Artificial Neural Network for fetus health

ANN is the cut edge between ML and DL [165]. As seen in Figure 4, ANN was used in (n=7, 6.50%) studies [69], [74], [78], [79], [83], [132], [133]. ANN models were used for the identification of growth diseases [74], [78], [79], fetus gender [83], facial expressions [133], face anatomical landmarks (Nasal Bone ROI) [132], and anatomical structures (nuchal translucency (NT)) [69]. Hence, ANN was not utilized to identify brain or heart structures, nor accompanying disease. Finally, we concluded that the first utilization of ANN to monitor fetus health was in 2010 [69]; the main goal was to detect NT, which helps early identify Down syndrome. The second utilization of ANN was completed in [2011] [132] to detect nasal bone. This detection also helps for early identification of Down syndrome.

### Reinforcement learning for fetus health

The first attempt to utilize RL to monitor fetus health was in 2019 [115]. Dueling Deep Q Networks (DDQN) was employed with CNN-ANN for objective terminated (LAT) to identify the standard planes (trans-thalamic (TT) and trans-cerebellar (TC)). The second time RL was utilized to monitor fetus health was reported in a recent study from 2021 [116]. A multi-agent RL (MARL) was used in a framework that compromises RL with multi-objective reinforcement learning (NAS) as well as gradient-based differentiable architecture sampler (GDAS). This framework achieved promising results for identifying the brain standard plane and localize mid-sagittal, transverse (T), coronal (C) planes in volumes, trans-thalamic (TT), trans-ventricular (TV), and trans-cerebellar (TC).

### 6.1.1 Strengths

To the best of our knowledge, this survey is the first to explore all AI techniques implemented to provide fetus health care and monitoring during different pregnancy trimesters. This survey may be considered comprehensive as it does not focus on specific clinical cases or diseases. Rather, it provides a holistic view of AI’s role in monitoring fetuses via ultrasound images. This survey will benefit both readers, first medical professionals to have a sight about the current AI technology in a development stage and assist data scientists in overcoming existing research and implication problems. secondly, for data scientists to continue future investigating in developing lightweight models and overcoming problems that slow the transformation of this technology based on a solid guideline proposed by the medical professionals. It may be considered a robust and high-quality survey. Because the most prominent health and information technology databases were searched using a well-developed search query, the search was sensitive and exact. Lastly, the utilization of techniques like scanning grey literature databases (Google Scholar), biomedical literature database (PubMed, Embase) World’s leading source citation databases (Web of Science), Subject Specific Databases (PsycINFO), and world’s leading source for scientific, technical, and medical research (ScienceDirect, IEEE explore, ACM Library) indicates that this study’s risk of publication bias is low.

### 6.1.2 Limitations

A comprehensive systematic survey was conducted. However, the search within the databases was conducted between 22nd and 23rd of June 2022, so we might miss some new studies. Besides the search terms "pregnant" or "pregnancy" or "uterus" were not used as search terms to identify relevant papers. Therefore, some fetus health monitoring studies might have been missed, lowering the overall number of studies. Additionally, this survey focused more on a clinical rather than the technical; therefore, some technical details may have been missed. Lastly, only English studies were included in the search. As a result, research written in other languages was excluded.

### 6.2 Practical and research implication

The diagnosis of US imaging plays a significant role in clinical patient care. Deep learning especially the CNN-based model, has lately gained much interest because of its excellent image recognition and performance. CNN were to live up to its potential in radiology, it is expected to aid sonographers and radiologists in achieving diagnostic perfection and improving patient care [166]. Diagnostic software based on artificial intelligence is vital in academic research and has significant media relevance. These systems are largely based on analyzing diagnostic images, such as X-RAY, CT, MRI, electronic horizons, and electroencephalograms. In contrast, US imaging is non-invasive, non-expensive, and non-ionizing, has limited AI applications compared to other radiology imaging technologies [167].

This survey found one available AI-based application to identify fetal cardiac substrucrures and indicate structural abnormalities as seen in [140]. From these findings, many studies achieved promising results in diagnosing fetus diseases or identifying specific fetus landmarks. However, to the best of our knowledge there is no randomized controlled trial (RCT) or pilot study carried out at a medical center or any adaptation of an AI-based application at any hospitals. This hesitance could be due to the following challenges [167]: 1) The present AI system can complete tasks that radiologists are capable of but may make mistakes that a radiologist would not make. For example, radiologists may not see undetectable modifications to the input data. These changes may not be seen by human eyes but would nevertheless impact the result of AI system’s categorization. In other words, a tiny variation might cause a deep learning system to reach a different result or judgment. 2) Developers require a specific quantity of reliable and standardized data with authorized reference standard to train AI systems. Annotated images may become an issue if this is done through retrospective research. Datasets may also be difficult to get as the firms that control them want to keep them private and preserve their intellectual property. Validation of an AI system in the clinical setting can be difficult as it frequently necessitates multistitutional collaboration and efficient communication between AI engineers and radiologists. Validating an AI system is also expensive and time-consuming. Finally, 3) when extensive patient databases are involved, ethical and legal concerns may arise.

This survey found some of these challenges were addressed in some studies. For example, applying transfer learning on AI model learned natively on US images and fine-tuning the model on an innovative data set collected from a different medical center and/or different US equipment. Another way to address the challenges of a small data set is to employ data augmentation (e.g., tissue resampling, future invadly being used in breast and chest imaging). Large quantities of medical images with strong reference standards are accessible in various fields, allowing the AI system to be trained. Other subspecialties like fetus disease, musculoskeletal disease or interventional radiology are less familiar with utilizing AI. However, it seems that in the future, AI applications may influence every medical application that utilizes any images [169].
This survey found that no standard guideline is being followed or developed that focused on AI in diagnostic imaging to meet clinical setting, evaluation, and requirements. Therefore, the Radiology Editorial Board [170] has created a list of nine important factors to assist in evaluating AI research as a first step. These recommendations are intended to enhance the validity and usefulness of AI research in diagnostic imaging. These issues are outlined for authors, although manuscript reviewers and readers may find them useful as well: 1) Carefully specify the AI experiment’s three image sets (training, validation, and test sets of images). 2) For final statistical reporting, use a separate test set. Overfitting is a problem with AI models, which means they only function for the images they are trained to recognize. It is ideal to utilize an outside collection of images (e.g., the external test set) from another center. 3) Use multivendor images for each step of the AI assessment, if possible (training, validation, test sets). Radiologists understand medical images from one vendor are not identical to theirs; radiomics and AI systems can help in identifying changes in the images. Moreover, multivendor AI algorithms are considerably more interesting than vendor-specific AI algorithms. 4) Always justify the training, validation, and test set sizes. Depending on the application, the number of images needed to train an AI system varies. After just a few hundred images, an AI model may be able to learn image segmentation. 5) Train the AI algorithm using a generally recognized industry standard of reference. For instance, chest radiographs are interpreted by a team of experienced radiologists. 6) Describe any image processing for the AI algorithm. Did the authors manually pick important images, or crop images to a limited field of view? How images are prepared and annotated has a big impact on radiologist’s comprehension of the AI model. 7) Compare AI performance against the expert radiology professionals. Competitions and leader boards for the “best” AI are popular among computer scientists working in the AI field. The area under the receiver operating characteristic curve is often used to compare one AI to another (AUC). On the other hand, physicians are considerably more interested in comparing the AI system to expert readers, when treating a patient. Experienced radiologist readers are recommended to benchmark an algorithm intended to identify radiologic anomalies. 8) Demonstrate the AI algorithm’s decision-making process. As previously mentioned, computer scientists working on imaging studies often describe their findings as a single AUC/ACC number: This AUC/ACC is compared to the previous best algorithm, which is a competitor. Unfortunately, the AUC/ACC value has little bearing on clinical medicine on its own. Even with a high AUC/ACC of 0.95, there may be an operating mode in which 99 out of 100 anomalies are overlooked. Many research teams overlay colored probability maps from the AI on the source images to assist doctors in comprehending the AI performance. 9) The AI algorithm should be open source so performance claims may be independently validated; this is already a prevalent recommendation, as this survey found only (n=9, 8.4%) studies [62, 70, 92, 105, 138–141, 163] made their works publicly available.

### 7. Future work and conclusion

Throughout this survey, various ways that AI techniques have been used to improve fetal care during pregnancy have been discussed, including: 1) determining the existence of a live embryo/fetal and estimate the pregnancy’s age, 2) identifying congenital fetus defects and determining the fetus’s location, 3) determining the placenta’s location, 4) checking for cervix opening or shortening and evaluating fetal development by measuring the quantity of amniotic fluid surrounding the baby, 5) evaluating the fetus’s health and see whether it is growing properly. Unfortunately, due to both ethical and privacy concerns as well as the reliability of AI decisions, all aspects of these models are not yet utilized as end applications in medical centers.

Undergraduate sonologist education is becoming more essential in medical student careers; consequently, new training techniques are necessary. As seen in other medical fields, AI-based applications and tools were employed for medical and health informatics students [171]. Moreover, Al-mobile-based applications were useful for medical students, clinicians, and allied health workers alike [172]. To the best of our knowledge, there are no Al-based mobile applications available to assist radiologist or sonographer students in medical college. Besides that, medical students may face many challenges during fetal US scanning [173]. Therefore, our future work will propose an Al- mobile-based application to help radiology or sonography students identify features in fetus US image and answer the following research questions: 1) How is the lightweight model efficient and accurate when localizing the fetal head, abdomen, femur, and thorax from the first to the third trimester? 2) How efficient and accurate the lightweight model will identify the fetal gestational age (GA) from the first to the third trimester? 3) How will inter-rater reliability tests validate the proposed model and compare the obtained result with experts and students using the intra-class correlation coefficient (ICC) [174]?

We have concluded that AI techniques utilized US images to monitor fetal health from different aspects through various GA. Out of 107 studies, DL was the most widely used model, followed by ML, ANN, and RL. These models were used to implement various tasks, including classification, segmentation, object detection, and regression. We found that even the most recent studies rely on the 2D US followed by 3D, and that 4D is rarely used. Furthermore, we found that most of the work was targeted the fetus head followed by the body, heart, face, and abdomen. We identified the lead institutes in this field and their research.

This survey discusses the availability of the dataset for each category and highlighted their promising result. In addition, we analyzed each study independently and provided observations for the future reader. AI optimization and feature extraction methods were reported for each study to highlight the unique contribution. Moreover, DL novel and unique works were highlighted for each category. The research and practical implications were discussed, and prospective research and clinical practitioner recommendations were provided. Finally, a future research direction has been proposed to answer the gap in this survey.

**Key Points**

- Artificial intelligence and machine learning methods can be successfully employed to optimize fetus care.
- Deep learning methods are the most common approach to monitor fetal health and the fetal head analysis is gaining traction.
- Future research should focus on employing these technologies and developing an actual application that can be used at medical colleges and centers.

**Supplementary data**

Multimedia Appendix 1: Search strings
Multimedia Appendix 2: Data extraction form

**Data Availability Statement**

Available upon request

**Funding**

No funding

**Conflicts of Interest**

None declared

**References**

[1] D. Hassabis, D. Kumaran, C. Summerfield, and M. Botvinick, “Neuroscience-Inspired Artificial Intelligence,” Neuron, vol. 95, no. 2, pp. 245–258, Jul. 2017, doi: 10.1016/j.neuron.2017.06.011.

[2] D. D. Miller and E. W. Brown, “Artificial Intelligence in Medical Practice: The Question to the Answer?,” Am. J. Med., vol. 131, no. 2, pp. 129–133, Feb. 2018, doi: 10.1016/j.amjmed.2017.10.035.

[3] P. Savadjiev et al., “Demystification of AI-driven medical image interpretation: past, present and future,” European Radiology, vol.
[29] S. P. Arjunan and M. C. Thomas, "A Review of Ultrasound Imaging Techniques for the Detection of Down Syndrome," *Irbm*, vol. 41, no. 2, pp. 115–123, 2020, doi: 10.1016/j.irbm.2019.10.004.

[30] J. Hartkopf, J. Moser, F. Schlegel, H. Preissl, and J. Keune, "Changes in event-related brain responses and habituation during child development – A systematic literature review," *Clin. Neurophysiol.*, vol. 130, no. 12, pp. 2238–2254, 2019, doi: 10.1016/j.clinph.2019.08.029.

[31] O. J. Shyne, J. A. P. Singh, and B. P. Shan, "A Review on techniques for computer aided diagnosis of soft markers for detection of down syndrome in ultrasound fetal images," *Biomed. Pharmacol. J.*, vol. 10, no. 3, pp. 1559–1568, 2017, doi: 10.13050/bpj/1266.

[32] J. Torrents-Barrera et al., "Segmentation and classification in MRI and US fetal imaging: Recent trends and future prospects," *Med. Image Anal.,* vol. 51, pp. 61–88, 2020, doi: 10.1016/j.media.2018.02.006.

[33] P. Kaur, G. Singh, and P. Kaur, "A Review of Denoising Medical Images Using Machine Learning Approaches," *Curr. Med. Imaging Rev.,* vol. 14, no. 5, pp. 675–685, 2017, doi: 10.2174/1573405166710428154156.

[34] L. Davidson and M. R. Boland, "Towards deep phenotyping pregnancy: a systematic review of artificial intelligence and machine learning methods to improve pregnancy outcomes," *Brief. Bioinform.,* 2021, doi: 10.1093/bib/bbaa369.

[35] S. Al-yousif et al., "A systematic review of automated preprocessing, feature extraction and classification of cardiotocography," *PeerJ Comput. Sci.,* vol. 7, pp. 1–37, 2021, doi: 10.7717/peerj-cs.452.

[36] M. Komatsu et al., "Towards clinical application of artificial intelligence in ultrasound imaging," *Biomedicines,* vol. 9, no. 7, 2021, doi: 10.3390/biomedicines9070720.

[37] W. M. Dos Santos, S. R. Secoli, and V. A. de A. Püschel, "The Joanna Briggs Institute approach for systematic reviews," *Rev. Lat. Am. Enfermagem,* vol. 26, p. e3074, 2018, doi: 10.1590/1518-8345.2885.3074.

[38] A. Liberati et al., "The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration," *J. Clin. Epidemiol.,* vol. 62, no. 10, pp. e1–e34, 2009, doi: 10.1016/j.jclinepi.2009.06.006.

[39] M. Ouzzani, H. Hammady, Z. Fedorowicz, and A. Elmagarmid, "Rayyan-a web and mobile app for systematic reviews," *Syst. Rev.,* vol. 5, no. 1, Dec. 2016, doi: 10.1186/s13643-016-0384-4.

[40] P. Kokol, H. Blázán Vojtěch, and J. Zavřenák, "Application of bibliometrics in medicine: a historical bibliometrics analysis," *Health Info. Libr. J.*, vol. 38, no. 2, pp. 125–138, Jun. 2021, doi: 10.1111/hir.12295.

[41] M. Bethune, E. Alibrahim, B. Davies, and E. Yong, "A pictorial guide for the second trimester ultrasound," *Australas. J. Ultrasound Med.*, vol. 16, no. 3, pp. 98–113, Aug. 2013, doi: 10.1002/j.2205-8834.2013.00014.x.

[42] Q. Huang and Z. Zeng, "A Review on Real-Time 3D Ultrasound Imaging Technology," *Biomed Res. Int.,* vol. 2017, 2017, doi: 10.1155/2017/602709.

[43] J. S. Abramowicz, "Nonmedical use of ultrasound: Bioeffects and safety risk," *Ultrasound Med. Biol.*, vol. 36, no. 8, pp. 1213–1220, 2010, doi: 10.1016/j.ultrasmedbio.2010.04.003.

[44] J. Abramowicz, "ALARA: The Clinical View," *Arch. Gynecol. Obstet.*, vol. 288, no. 4, pp. 2556–2562, Jun. 2009, doi: 10.1002/j.2205-8834.2009.e34.

[45] D. A. Driscoll and S. Gross, "Prenatal Screening for Aneuploidy," *N. Engl. J. Med.*, vol. 360, no. 24, pp. 2556–2562, Jun. 2009, doi: 10.1056/nejmsr0900134.

[46] Q. Huang and Z. Zeng, "A Review on Real-Time 3D Ultrasound Imaging Technology," *Biomed Res. Int.,* vol. 2017, 2017, doi: 10.1155/2017/602709.

[47] J. S. Abramowicz, "Nonmedical use of ultrasound: Bioeffects and safety risk," *Ultrasound Med. Biol.*, vol. 36, no. 8, pp. 1213–1220, 2010, doi: 10.1016/j.ultrasmedbio.2010.04.003.

[48] J. Abramowicz, "ALARA: The Clinical View," *Arch. Gynecol. Obstet.*, vol. 288, no. 4, pp. 2556–2562, Jun. 2009, doi: 10.1002/j.2205-8834.2010.00014.x.

[49] D. A. Driscoll and S. Gross, "Prenatal Screening for Aneuploidy," *N. Engl. J. Med.*, vol. 360, no. 24, pp. 2556–2562, Jun. 2009, doi: 10.1056/nejmsr0900134.

[50] Q. Huang and Z. Zeng, "A Review on Real-Time 3D Ultrasound Imaging Technology," *Biomed Res. Int.,* vol. 2017, 2017, doi: 10.1155/2017/602709.

[51] J. S. Abramowicz, "Nonmedical use of ultrasound: Bioeffects and safety risk," *Ultrasound Med. Biol.*, vol. 36, no. 8, pp. 1213–1220, 2010, doi: 10.1016/j.ultrasmedbio.2010.04.003.

[52] J. Abramowicz, "ALARA: The Clinical View," *Arch. Gynecol. Obstet.*, vol. 288, no. 4, pp. 2556–2562, Jun. 2009, doi: 10.1002/j.2205-8834.2010.00014.x.

[53] D. A. Driscoll and S. Gross, "Prenatal Screening for Aneuploidy," *N. Engl. J. Med.*, vol. 360, no. 24, pp. 2556–2562, Jun. 2009, doi: 10.1056/nejmsr0900134.

[54] Q. Huang and Z. Zeng, "A Review on Real-Time 3D Ultrasound Imaging Technology," *Biomed Res. Int.,* vol. 2017, 2017, doi: 10.1155/2017/602709.

[55] J. S. Abramowicz, "Nonmedical use of ultrasound: Bioeffects and safety risk," *Ultrasound Med. Biol.*, vol. 36, no. 8, pp. 1213–1220, 2010, doi: 10.1016/j.ultrasmedbio.2010.04.003.
diagnosis using variational level set method and classification using artificial neural networks,” in *Proceedings - 2014 5th International Conference on Signal and Image Processing, ICSP 2014*, 2014, pp. 303–309, doi: 10.1109/ICSP.2014.54.

[80] M. A. Maraci et al., “Toward point-of-care ultrasound estimation of fetal gestational age from the trans-cerebellar diameter using CNV-based ultrasound image analysis,” *J. Med. Imaging*, vol. 7, no. 1, p. 1, 2020, doi: 10.1117/1.jmi.7.1.014501.

[81] P. Chen et al., “A preliminary study to quantitatively evaluate the development of maturation degree for fetal lung based on transfer learning deep model from ultrasound images,” *Int. J. Comput. Assist. Radiol. Surg.*, vol. 15, no. 8, pp. 1407–1415, Aug. 2020, doi: 10.1007/s11548-020-02211-1.

[82] J. C. Prieto et al., “An automated framework for image classification and segmentation of fetal ultrasound images for gestational age estimation,” *Proc. SPIE - the Int. Soc. Opt. Eng.*, vol. 11596, p. 55, Feb. 2021, doi: 10.1117/12.2582243.

[83] I. M. D. Mayanjiaya, H. A. Nugroho, and N. A. Setiawan, “The classification of fetus gender on ultrasound images using learning vector quantization (LVQ),” in *Proceedings - 2014 Makassar International Conference on Electrical Engineering and Informatics, MICIEEE 2014*, 2014, pp. 150–155. doi: 10.1109/MICIEEE.2014.7067329.

[84] S. Nie, J. Yu, P. Chen, J. Zhang, and Y. Wang, “A novel method with a deep network and directional edges for automatic detection of a fetal head,” in *2015 23rd European Signal Processing Conference, EUSIPCO 2015*, 2015, pp. 654–658. doi: 10.1109/EUSIPCO.2015.7362464.

[85] J. Li et al., “Automatic Fetal Head Circumference Measurement in Ultrasound Using Random Forest and Fast Ellipse Fitting,” *IEEE J. Biomed. Health Informatics*, vol. 22, no. 1, pp. 215–223, 2018, doi: 10.1109/JBHII.2017.2703890.

[86] S. Nie, J. Yu, Y. Wang, J. Zhang, and P. Chen, “Shape model and marginal space of 3D ultrasound volume data for automatically detecting a fetal head,” in *ICALIP 2014 - 2014 International Conference on Audio, Language and Image Processing*, Proceedings, 2015, pp. 681–685. doi: 10.1109/ICALIP.2014.7009881.

[87] Z. Sobhaninia et al., “Fetal Ultrasound Image Segmentation for Measuring Biometric Parameters Using Multi-Task Deep Learning,” in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2019, pp. 6545–6548. doi: 10.1109/EMBC.2019.8856981.

[88] D. Qiao and F. Zulkernine, “Dilated Squeeze-and-Excitation U-Net for Fetal Ultrasound Image Segmentation,” in *2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology, CIBCB 2020*, 2020, pp. 1–7. doi: 10.1109/CIBCB48159.2020.9277667.

[89] Z. Sobhaninia et al., “Fetal Ultrasound Image Segmentation for Measuring Biometric Parameters Using Multi-Task Deep Learning,” in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2019, pp. 6545–6548. doi: 10.1109/EMBC.2019.8856981.

[90] P. Li, H. Zhao, P. Liu, and F. Cao, “Automated measurement network for accurate segmentation and parameter estimation in fetal head ultrasound images,” *Med. Biol. Eng. Comput.*, vol. 58, no. 11, pp. 2879–2892, Nov. 2020, doi: 10.1109/111517-020-02242-5.

[91] R. Droste, P. Chatelain, L. Drukker, H. Sharma, A. T. Papageorghiou, and J. A. Noble, “Discovering Salient Anatomical Landmarks by Predicting Human Gaze,” in *Proceedings - International Symposium on Biomedical Imaging*, 2020, vol. 2020-April, pp. 717–717. doi: 10.1109/ISBI45749.2020.9098505.

[92] B. Al-Bander, T. Alzahrani, S. Alzahrani, B. M. Williams, and Y. Zheng, “Improving Fetal Head Contour Detection by Object Localisation with Deep Learning,” in *Communications in Computer and Information Science, 2020*, vol. 1065 CCIS, pp. 142–150. doi: 10.1007/978-3-030-39343-4_12.

[93] M. C. Fiorentino, S. Moccia, M. Capparuccia, S. Giamberini, and E. Frontoni, “A regression framework for fetal-head-circumference delineation from US fetal images,” *Comput. Methods Programs Biomed.*, vol. 198, p. 105771, Jan. 2021, doi: 10.1016/j.cmpb.2020.105771.

[94] S. Fathimuthu Joharah and K. Mohideen, “Automatic detection of fetal ultrasound image using multi-task deep learning,” *J. Crit. Rev.*, vol. 7, no. 7, pp. 987–993, 2020, doi: 10.3183/cr.07.07.181.

[95] R. Qu, G. Xu, C. Ding, W. Jia, and M. Sun, “Deep Learning-Based Methodology for Recognition of Fetal Brain Standard Scan Planes in 2D Ultrasound Images,” *IEEE Access*, vol. 8, pp. 44443–44451, 2020, doi: 10.1109/ACCESS.2019.2950387.

[96] R. Qu, G. Xu, C. Ding, W. Jia, and M. Sun, “Standard plane identification in fetal brain ultrasound scans using a differential convolutional neural network,” *IEEE Access*, vol. 8, pp. 83821–83830, 2020, doi: 10.1109/ACCESS.2020.2991845.

[97] V. Singh et al., “Semantic Segmentation of Cerebellum in 2D Fetal Ultrasound Brain Images Using Convolutional Neural Networks,” *IEEE Access*, vol. 9, pp. 85864–85873, 2021, doi: 10.1109/ACCESS.2021.3088946.

[98] S. Wang et al., “Deep Learning based Fetal MIDDLE Cerebral Artery Segmentation in Large-scale Ultrasound Images,” in *Proceedings - 2018 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2018*, 2019, pp. 532–539. doi: 10.1109/BIBM.2018.8621510.

[99] M. Yaqub et al., “Volumetric segmentation of key fetal brain structures in 3D ultrasound,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8184, 2020.
C. Chen et al., “Region Proposal Network with Graph Prior and Iou-Balance Loss for Landmark Detection in 3D Ultrasound,” in Proceedings - International Symposium on Biomedical Imaging, 2020, vol. 2020-April, pp. 1829–1833. doi: 10.1109/ISBI45749.2020.9098368.

T. A. Anjit and S. Rishidas, “Identification of nasal bone for the early detection of down syndrome using Back Propagation Neural Network,” in ICCSP 2011 - 2011 International Conference on Communications and Signal Processing, 2011, pp. 136–140. doi: 10.1109/ICCSP.2011.5739286.

P. R. Dave and M. S. B. Nadid, “Facial expressions extraction from 3D sonography images,” in Proceedings of 2015 IEEE International Conference on Electrical, Computer and Communication Technologies, ICECCT 2015, 2015, pp. 1–6. doi: 10.1109/ICECCT.2015.7266050.

Y. Miyagi, T. Hata, S. Bouno, A. Koyanagi, and T. Miyake, “Recognition of facial expression of fetuses by artificial intelligence (AI),” J. Perinat. Med., vol. 49, no. 5, pp. 596–603, Jun. 2021, doi: 10.1515/jpm-2020-0537.

M. Dinesh Simon and A. R. Kavitha, “Ultrasound Detection of Down Syndrome Using Multiscale Quantiﬁer with Convolutional Neural Network,” in Computational Optimization Techniques and Applications, IntechOpen, 2021. doi: 10.5772/intechopen.95602.

J. Tan et al., “Automated Detection of Congenital Heart Disease in Fetal Ultrasound Screening,” in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 12437 LNCS, Springer, 2020, pp. 243–252. doi: 10.1007/978-3-030-60334-2_24.

T. Yang et al., “Segmentation of Five Components in Four Chamber View of Fetal Echocardiography,” in Proceedings - International Symposium on Biomedical Imaging, 2020, vol. 2020-April, pp. 1962–1965. doi: 10.1109/ISBI45749.2020.9098726.

R. Arnaout, L. Curran, Y. Zhao, J. C. Levine, E. Chinn, and A. J. Moon-Grady, “An ensemble of neural networks provides expert-level prenatal detection of complex congenital heart disease,” Nat. Med., vol. 27, no. 5, pp. 882–891, 2021, doi: 10.1038/s41591-021-01342-5.

Y. Gong et al., “Fetal congenital heart disease echocardiogram screening based on dgacman: Adversarial one-class classiﬁcation combined with video transfer learning,” IEEE Trans. Med. Imaging, vol. 39, no. 4, pp. 1206–1222, 2020, doi: 10.1109/TMI.2019.2946059.

M. Komatsu et al., “Detection of cardiac structural abnormalities in fetal ultrasound video using deep learning,” Appl. Sci., vol. 11, no. 1, pp. 1–12, 2021, doi: 10.3390/app11010371.

A. Dozen et al., “Image segmentation of the ventricular septum in fetal cardiac ultrasound videos based on deep learning using time-series information,” Biomolecules, vol. 10, no. 11, pp. 1–17, 2020, doi: 10.3390/biom10111526.

A. Patra, W. Huang, and J. A. Noble, “Learning spatio-temporal aggregation for fetal heart analysis in ultrasound videos,” in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 10553 LNCS, Springer, 2017, pp. 276–284. doi: 10.1007/978-3-319-67558-9_32.

L. Xu, M. Liu, J. Zhang, and Y. He, “Convolutional-Neural-Network-Based Approach for Segmentation of Apical Four-Chamber View from Fetal Echocardiography,” IEEE Access, vol. 8, pp. 80347–80406, 2020, doi: 10.1109/ACCESS.2020.2984630.

L. Xu et al., “DW-Net: A cascaded convolutional neural network for apical-four-chamber view segmentation in fetal echocardiography,” Comput. Med. Imaging Graph., vol. 80, p. 101690, Mar. 2020, doi: 10.1016/j.compmedimag.2019.101690.

V. Sundaresh, C. P. Bridge, C. Ioannou, and J. A. Noble, “Automated characterization of the fetal heart in ultrasound images using fully convolutional neural networks,” in Proceedings - International Symposium on Biomedical Imaging, 2017, pp. 671–674. doi: 10.1109/ISBI.2017.7950609.

J. Dong et al., “A Generic Quality Control Framework for Fetal Ultrasound Cardiac Four-Chamber Planes,” IEEE J. Biomed. Heal. Informatics, vol. 24, no. 4, pp. 931–942, 2020, doi: 10.1109/JBHI.2019.2948316.

B. Pu, N. Zhu, K. Li, and S. Li, “Fetal cardiac cycle detection in multi-resource echocardiograms using hybrid classiﬁcation framework,” Futur. Gener. Comput. Syst., vol. 115, pp. 825–836, 2021, doi: 10.1016/j.future.2020.09.014.

B. Rahmatullah, I. Sarris, A. Papageorgiou, and J. A. Noble, “Quality control of fetal ultrasound images: Detection of abnormal anatomical landmarks using AdaBoost,” in Proceedings -
International Symposium on Biomedical Imaging, 2011, pp. 6–9. doi: 10.1109/ISBI.2011.5872342.

[149] X. Yang et al., “Standard plane localization in ultrasound by radial component,” in 2014 IEEE 11th International Symposium on Biomedical Imaging, ISBI 2014, 2014, pp. 1180–1183. doi: 10.1109/Isbi.2014.6868086.

[150] L. Wu, J. Z. Cheng, S. Li, B. Lei, T. Wang, and D. Ni, “FUlQA: Fetal ultrasound image quality assessment with deep convolutional networks,” IEEE Trans. Cybern., vol. 47, no. 5, pp. 1336–1349, 2017, doi: 10.1109/TCYB.2017.2671898.

[151] D. Ni et al., “Standard Plane Localization in Ultrasound by Radial Component Model and Selective Search,” Ultrasound Med. Biol., vol. 40, no. 11, pp. 2728–2742, Nov. 2014, doi: 10.1016/j.ultrasmedbio.2014.06.006.

[152] B. Rahmatullah, A. T. Papageorghiou, and J. Alison Noble, “Image analysis using machine learning: Anatomical landmarks detection in fetal ultrasound images,” in Proceedings - International Computer Software and Applications Conference, 2012, pp. 354–355. doi: 10.1109/COMPSAC.2012.52.

[153] H. Chen, D. Ni, X. Yang, S. Li, and P. A. Heng, “Fetal Abdominal Standard Plane Localization Through Representation Learning with Knowledge Transfer,” in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 8679, 2014, pp. 125–132. doi: 10.1007/978-3-319-10581-9_16.

[154] B. Rahmatullah, A. Papageorghiou, and J. A. Noble, “Automated selection of standardized planes from ultrasound volume,” in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 7009 LNCs, 2011, pp. 35–42. doi: 10.1007/978-3-642-24319-6_5.

[155] B. Kim, K. C. Kim, Y. Park, J. Y. Kwon, J. Jang, and J. K. Seo, “Machine-learning-based automatic identification of fetal abdominal circumference from ultrasound images,” Physiol. Meas., vol. 39, no. 10, p. 105007, 2018, doi: 10.1088/1361-6579/aaac25.

[156] P. Deepika, R. M. Suresh, and P. Pabitha, “Defending Against Child Death: Deep learning-based diagnosis method for abnormal identification of fetus ultrasound images,” Comput. Intell., vol. 37, no. 1, pp. 128–154, 2021, doi: 10.1111/coin.12394.

[157] J. Jang, Y. Park, B. Kim, S. M. Lee, J. Y. Kwon, and J. K. Seo, “Automatic Estimation of Fetal Abdominal Circumference from Ultrasound Images,” IEEE J. Biomed. Heal. Informatics, vol. 22, no. 5, pp. 1512–1520, Nov. 2017, doi: 10.1109/JBHI.2017.2776116.

[158] B. C. Loweckamp, D. T. Chen, L. Ibáñez, and D. Blezek, “The design of simpleITK,” Front. Neuroinformat., vol. 7, no. DEC, p. 45, Dec. 2013, doi: 10.3389/fninf.2013.00045.

[159] P. Yushkevich, J. Piven, H. Cody, and S. Ho, “User-guided level set segmentation of anatomical structures with ITK-SNAP,” Neuroimage, vol. 31, no. 3, pp. 1116–1128, 2006, Accessed: Oct. 03, 2021. [Online]. Available: https://www.academia.edu/download/41218070/User-Guided_Level_Set_Segmentation_of_An20160113-9594-xfzom3.pdf20160115-19908-odelqv.pdf

[160] T. L. a. van den Heuvel, D. de Bruijn, C. I. de Korte, and B. van Ginneken, “Automated measurement of fetal head circumference using 2D ultrasound images,” PLoS One, vol. 13, no. 8, Aug. 2018, doi: 10.1371/journal.pone.0200412.

[161] J. Zhang, C. Petitjean, P. Lopez, and S. Ainouz, “Direct estimation of fetal head circumference from ultrasound images based on regression CNN,” in Proceedings of Machine Learning Research, 2020, vol. 121, pp. 914–922. [Online]. Available: http://proceedings.mlr.press/v121/zhang20a.html

[162] S. Pieper, M. Halle, and R. Kikinis, “3D Slicer,” in 2004 2nd IEEE International Symposium on Biomedical Imaging: Macro to Nano, 2004, vol. 1, no. 3, pp. 632–635. doi: 10.1007/s00133-019-0654-4.

[163] X. Yang et al., “Hybrid attention for automatic segmentation of whole fetal head in prenatal ultrasound volumes,” Comput. Methods Programs Biomed., vol. 194, p. 105519, 2020, doi: 10.1016/j.cmpb.2020.105519.

[164] C. F. Baumgartner et al., “SonoNet: Real-Time Detection and Localisation of Fetal Standard Scan Planes in Freehand Ultrasound,” IEEE Trans. Med. Imaging, vol. 36, no. 11, pp. 2204–2215, Nov. 2017, doi: 10.1109/TMI.2017.2712367.

[165] N. K. Chauhan and K. Singh, “A review on conventional machine learning vs deep learning,” in 2018 International Conference on Computing, Power and Communication Technologies, GUCON 2018, Mar. 2019, pp. 347–352. doi: 10.1109/GUCON.2018.8675097.

K. Yasaka and O. Abe, “Deep learning and artificial intelligence in radiology: Current applications and future directions,” PloS Med., vol. 15, no. 11, Nov. 2018, doi: 10.1371/journal.pmed.1002707.

B. Shuo, Wang, M. Ji-Bin, Liu, M. Ziyin, Zhu, and P. John, Eisenbrey, “Artificial Intelligence in Ultrasound Imaging: Current Research and Applications,” Adv. ULTRASOUND DIAGNOSIS Ther., vol. 3, no. 3, p. 53, Sep. 2019, doi: 10.37015/audt.2019.190811.

Z. Akkus et al., “A Survey of Deep-Learning Applications in Ultrasound: Artificial Intelligence–Powered Ultrasound for Improving Clinical Workflow,” J. Am. Coll. Radiol., vol. 16, no. 9, pp. 1318–1328, 2019, doi: 10.1016/j.acra.2019.06.004.

E. Neri et al., “What the radiologist should know about artificial intelligence – an ESR white paper,” Insights Imaging, vol. 10, no. 1, Dec. 2019, doi: 10.1186/s13244-019-0738-2.

D. A. Bluemke et al., “Assessing radiology research on artificial intelligence: A brief guide for authors, reviewers, and readers-from the Radiology Editorial Board,” Radiology, vol. 294, no. 2, pp. 487–489, 2020, doi: 10.1148/radiol.2019192515.

A. Hasan Sapci and H. Aylin Sapci, “Artificial intelligence education and tools for medical and health informatics students: Systematic review,” JMR Med. Educ., vol. 6, no. 1, 2020, doi: 10.2196/19285.

I. M. Fires, G. Marques, N. M. Garcia, F. Flórez-revuelta, V. Ponciano, and S. Oniani, “A research on the classification and applicability of the mobile health applications,” J. Pers. Med., vol. 10, no. 1, 2020, doi: 10.3390/jpm10010011.

D. P. Bahner, A. Jasne, S. Boone, A. Mueller, and E. Cortez, “The ultrasound challenge a novel approach to medical student ultrasound education,” J. Ultrasound Med., vol. 31, no. 12, pp. 2013–2016, Dec. 2012, doi: 10.7863/jum.2012.31.12.13.

J. J. Bartko, “The intraclass correlation coefficient as a measure of reliability,” Psychol. Rep., vol. 19, no. 1, pp. 3–11, Sep. 1966, doi: 10.2466/pr0.1966.19.1.3.