FACULDADE DE MEDICINA DA UNIVERSIDADE DE COIMBRA

MESTRADO INTEGRADO EM MEDICINA – TRABALHO FINAL

CARLA RAQUEL DA SILVA BARROS NETO

Application of Artificial Intelligence Models In Second Trimester Obstetric Ultrasonography

ARTIGO DE REVISÃO NARRATIVA

ÁREA CIENTÍFICA DE OBSTETRÍCIA

Trabalho realizado sob a orientação de:
PROFESSORA DOUTORA ANA LUÍSA FIALHO AMARAL DE AREIA
DRA IOLANDA JOÃO MORA CRUZ DE FREITAS FERREIRA

ABRIL/2023

Application of Artificial Intelligence Models In Second Trimester Obstetric Ultrasonography

Carla Raquel da Silva Barros Neto[1], Iolanda João Mora Cruz De Freitas Ferreira[1,2], Ana Luísa Fialho Amaral de Areia[1,2]

[1]Faculdade de Medicina, Universidade de Coimbra (FMUC), Portugal

[2]Maternidade Doutor Daniel de Matos, Centro Hospitalar e Universitário de Coimbra, EPE, Coimbra, Portugal

Centro Hospitalar e Universitário de Coimbra, EPE - Maternidade Doutor Daniel de Matos - Rua Miguel Torga 1, 3000-157 Coimbra, Portugal

Endereço de correio eletrónico: carlaraquel1997@gmail.com

Index

List of abbreviations and acronyms

AC – Abdominal Circumference

AFI – Amniotic Fluid Index

AI – Artificial Intelligence

AIUM – American Institute of Ultrasound in Medicine

AUC – Area Under the Curve

BOD – Bilateral Orbital Diameter

BPD – Biparietal Diameter

CHD – Congenital Heart Defects

CNN – Convolutional Neural Networks

CNS – Central Nervous System

DGS – Direção Geral da Saúde

DL – Deep Learning

EFW – Estimated Fetal Weight

FL – Femur Length

GA – Gestational Age

GAN – Generative Adversarial Networks

HC – Head Circumference

IOD – Interorbital Diameter

ISUOG- International Society of Ultrasound in Obstetrics and Gynecology

ML – Machine Learning

OFD – Occipitofrontal Diameter

US – Ultrasound

VD – Distance Between Vertex and Nasion

3VT – Three-vessel Trachea View

Abstract

Introduction: Second trimester ultrasound is essential for the assessment of fetal anatomy, growth and well-being, and is therefore a time-consuming examination requiring multiple tasks to be performed. Although there are many advantages to its use in obstetrics, it is a technique with inherent limitations and is a highly operator-dependent diagnostic method, and there may be intra- and inter-observer variability. Therefore, the results obtained may sometimes be lower than expected. The emergence of artificial intelligence technology may be the way to overcome these limitations by being able to process large amounts of data to identify patterns or using deep learning algorithms. These include convolutional neural network algorithms, which are most widely used in medical imaging. Through this narrative review, we aim to understand how the application of artificial intelligence models can contribute to improve the diagnostic sensitivity and workflow of second trimester ultrasound.

Methodology: A literature search was performed in Pubmed, Elsevier and WebOfScience databases using the MeSH terms: "artificial intelligence" and "ultrasonography". Articles published between 2012 and 2022, in English, Portuguese or Spanish, that addressed the application of artificial intelligence models in second trimester obstetric ultrasound were selected.

Discussion: Fetal anatomical evaluation is essential during the second trimester to diagnose congenital anomalies, taking timely action and counselling. Congenital heart defects are the most common anomaly, but only 30% of cases are diagnosed by ultrasound. Artificial intelligence models have proven to be effective in the segmentation of structures and detection of diagnostic planes at various anatomical points recommended to be assessed by international guidelines. Regarding the heart and central nervous system, its application has shown to obtain values that correspond to a high diagnostic capacity of these congenital anomalies. Second trimester ultrasound also allows gestational age and fetal biometry calculation. This can diagnose growth anomalies and estimate the gestational age if it has not been done earlier in pregnancy. Automation of fetal biometry resulted in lower measurement errors than those obtained by specialists, which also translated into more correct gestational age estimates compared to those observed in clinical practice in the second trimester. In order to overcome the subjectivity inherent in the qualitative assessment of these structures, automatic evaluation of the placenta and amniotic fluid has also shown promising results in some models. By automating workflow analysis, it is possible to understand the needs of ultrasound operators and create new ways of performing the exam to achieve better results.

Conclusion: The application of artificial intelligence models shows promising results in diagnostic improvement in second trimester ultrasound. It would be important to improve and develop better models specifically for fetal anomaly detection in this trimester of pregnancy to improve their diagnostic ability. Moreover, further prospective studies to adapt the models and confirm their clinical applicability and potential advantages in daily practice are very important.

Keywords: Artificial intelligence; deep learning; ultrasonography; pregnancy trimester, second; congenital malformations.

Resumo

Introdução: A ecografia do segundo trimestre é essencial para a avaliação da anatomia fetal, crescimento e bem-estar, sendo por isso um exame moroso que exige a realização de múltiplas tarefas. Embora haja muitas vantagens na sua utilização em obstetrícia, trata-se de uma técnica com limitações intrínsecas e constitui um método de diagnóstico altamente dependente do operador, podendo haver variabilidade intra e interobservador. Consequentemente, os resultados obtidos podem por vezes ser inferiores ao esperado. A emergência da tecnologia de inteligência artificial pode ser a forma de ultrapassar estas limitações, ao ser capaz de processar grandes quantidades de dados e identificar padrões, ou utilizar algoritmos de aprendizagem profunda. Estes incluem o algoritmo de rede neural convolucional, que é o mais amplamente utilizado em imagiologia médica. Através desta revisão narrativa, pretendemos compreender como a aplicação de modelos de inteligência artificial pode contribuir para melhorar a sensibilidade diagnóstica e o fluxo de trabalho da ecografia do segundo trimestre.

Metodologia: A pesquisa bibliográfica foi realizada nas bases de dados *Pubmed*, *Elsevier* e *WebOfScience* usando os termos *MeSH*: "artificial intelligence" e "ultrasonography". Foram selecionados artigos publicados entre 2012 e 2022, em Inglês, Português ou Espanhol, que abordavam a aplicação de modelos de inteligência artificial na ecografia obstétrica do segundo trimestre.

Discussão: A avaliação anatómica fetal é essencial durante o segundo trimestre para diagnosticar anomalias congénitas, e assim tomar medidas e aconselhamento atempados. As cardiopatias congénitas são as mais frequentes, porém a ecografia apenas diagnostica 30% dos casos. Os modelos de inteligência artificial provaram ser eficazes na segmentação de estruturas e na deteção de planos de diagnóstico em vários pontos anatómicos recomendados para serem avaliados pelas *guidelines* internacionais. Relativamente ao coração e ao sistema nervoso central, a sua aplicação demonstrou obter valores que correspondem estatisticamente a uma elevada capacidade diagnóstica de anomalias congénitas nestes órgãos e sistemas. A ecografia do segundo trimestre também permite o cálculo da idade gestacional e da biometria fetal. Assim, é possível diagnosticar anomalias de crescimento e estimar a idade gestacional se esta não tiver sido realizada mais cedo na gravidez. A automatização da biometria fetal revelou erros de medição inferiores aos obtidos por especialistas, o que também se traduz na obtenção de estimativas da idade gestacional mais corretas comparadas com as observadas na prática clínica neste trimestre. De modo a superar a subjetividade inerente à avaliação qualitativa destas estruturas, a avaliação automática da placenta e do líquido amniótico também mostrou resultados promissores em alguns modelos. A automatização da análise do fluxo de trabalho permite compreender as necessidades dos ecografistas e, assim, criar formas da realização do exame que permitam obter melhores resultados.

Conclusão: A aplicação de modelos de inteligência artificial apresenta resultados promissores na melhoria da sensibilidade diagnóstica na ecografia do segundo trimestre. Seria importante aperfeiçoar e desenvolver modelos específicos para a deteção de anomalias fetais neste trimestre da gravidez. Além disso, são muito importantes novos estudos prospetivos para adaptar os modelos e confirmar a sua aplicabilidade clínica e potenciais vantagens na prática diária.

Palavras-chave: Inteligência artificial; aprendizagem computacional; ecografia; segundo trimestre da gravidez; malformações congénitas.

1. Introduction

1.1 Obstetric Ultrasound

Ultrasound (US) was first used for therapeutic purposes in the 1920s, taking advantage of its thermal effects. It was first used as a diagnostic tool in obstetrics by Professor Ian Donald and his team, who published the first paper on the subject in 1958. Professor Ian Donald also recorded the first US image of a fetus in the 1980s. It was taken in bright mode (white and black dots) and corresponded to the head of the fetus.[1]

Because it is non-invasive, does not use ionizing radiation, and does not interfere with the development or growth of the fetus, it is not harmful to either the mother or the fetus. It is also inexpensive and provides an immediate image.[2] Despite its many advantages, US is still an operator-dependent task that requires extensive training programs for physicians. In addition to the disadvantages inherent in the technique itself, characteristics such as the size and movements of the fetus and the amount of adipose tissue in the mother contribute to the difficulty in obtaining quality images. For these reasons, US results can sometimes be less than expected.[3–5] In an attempt to standardize the performance of obstetric US scans, the International Society of Ultrasound in Obstetrics and Gynecology (ISUOG) has developed several practice guidelines that explain what they consider to be best practice.[6]

1.2 Mid-trimester Ultrasound

Fetal assessment requires the acquisition, identification, and segmentation of US images. These are time-consuming processes that require specific knowledge and generally need to be performed by trained and specialized sonographers. For this reason, US is a technique that requires the existence of guidelines and protocols to achieve better results. The aim is to provide parents with valid prenatal information and to define the best therapeutic approach before, during or after birth.[3]

ISUOG guidelines have defined the US criteria that should be evaluated in the mid-trimester, between 18 and 24 weeks, and it has been recommended by the Portuguese Health Authority (DGS) that this US should be performed between 20 and 22 weeks.[6,7] In addition to assessing congenital anatomical defects, especially those that may not have been observed in the first trimester, mid-trimester US also evaluates other pregnancy structures (amniotic fluid and placenta), fetal biometry, which assesses fetal growth, and fetal well-being. Fetal growth monitoring is important for comparison with later US scans.[6] Taking all these parameters into account, a complete mid-trimester US scan requires the acquisition of more than 20 different images and planes.[8]

According to the American Institute of Ultrasound in Medicine (AIUM), in a high-risk pregnancy, routine US should be supplemented by more specific evaluation, especially in the last two trimesters of pregnancy. This is because risk factors such as a history of a previous abnormal fetus, gestational diabetes, advanced maternal age, increased body mass index or use of teratogenic drugs increase the likelihood of fetal anatomical, genetic or growth abnormalities in the current pregnancy.[9]

The Eurofetus study was developed to understand the diagnostic ability of obstetric US. With the sample used, they concluded that the overall diagnostic sensitivity for fetal anomalies was about 56%, with only 61% diagnosed in utero. Furthermore, it should be noted that the diagnostic sensitivity of US is higher for major anomalies when comparing with minor ones.[10]

One of the most important evaluated structures on the second trimester scan is the heart. Congenital heart defects (CHD) are the most frequent abnormalities, occurring in about 9 of 1000 births. They are responsible for a first-year mortality rate of about 18%.[11] Despite this, the ability to diagnose major and minor fetal heart defects by US is only 38.8% and 20.8%, respectively. Congenital central nervous system (CNS) anomalies are also frequent. In this case, the diagnostic sensitivity of obstetric US is higher, but it remains undiagnosed in almost 12% of cases.[10]

Fetal biometry, or fetal growth assessment, is performed by taking several measurements, namely: biparietal diameter (BPD), head circumference (HC), abdominal circumference (AC) and femur length (FL). In order to obtain these correctly, assurance that the best plane and calliper placement are acquired is vital.[6] Three of these measurements (HC, AC and FL) are taken into account in the Hadlock formula, which is used to calculate the estimated fetal weight.[6,12] In the first trimester, crown-rump length measurement allows the correct determination of gestational age (GA), specifically when this parameter measures between 45 and 84 milimeters.[12] However, if we focus on at-risk populations that have less access to health care, GA may only be determined later in pregnancy. The fetal variability inherent to pregnancy development implies less accuracy in dating during these gestational ages.[13] Thus, guidelines indicate that, after the first trimester, GA can be inferred by measuring HC.[6] Therefore, more accurate GA calculation may be possible at this point in pregnancy if the mechanism for obtaining these measurements is automated.[3,4]

The placenta is the structure that ensures fetal development throughout pregnancy. It´s correct evaluation is currently very subjective because it relies not only on a quantitative, but also on a qualitative evaluation, namely regarding the evaluation of its morphology.[14] The second trimester US also assesses the relation between the placenta and the internal cervical os. According to the ISUOG guidelines, the distance between the internal cervical os and the lower limit of the placenta should be more than 15 millimetres. If this is not the case, a follow-up examination in the third trimester is justified. In addition, if there is a diagnosis of low-lying placenta or placenta previa, one should assess the presence or absence of ultrasonographic signals for placenta accreta.[6] The development of methods that are less dependent on the subjectivity inherent in current clinical practice, might improve assessment of placental integrity, maturity, and location.[14]

1.3 Artificial Intelligence

Since US is a technique still very dependent on the manual function of the operators and dependent on the analysis done by them, the automation of these processes, achieved through artificial intelligence models, may be the solution to obtain better results.[15]

Artificial intelligence (AI) is a technology that attempts to mimic functions performed by the human brain using a computer program.[15] Research into computer models that attempted to mimic human reasoning began in the last century and was defined in the 1950s as “Intelligent Machinery”.[16] The goal of this software is to be taught and trained to perform a function with real data. After the development and application of an AI software, it can be called a model, application, or algorithm.[15]

Data science enables the selection of the most meaningful data from a large volume of data for subsequent evaluation. This is usually applied to big data, which corresponds to large amounts of categorized data that are used in learning by an AI model. This ensures that the developed models are of high quality.[15] Nowadays, thanks to the computerization that has occurred in Medicine, there is a large amount of data, both clinical and imaging, stored in databases. These can be used as big data to help create robust automatic models that can achieve good results in clinical practice.[17]

Machine Learning (ML) is a subset of AI. It comprises the development of computer systems, which can learn without following explicit instructions, by using algorithms, and statistical models that analyse and draw inferences from patterns in data. These algorithms are able to draw these inferences from data with or without human aid or human data categorization.[15,16] It´s performance depends on the patterns that the machine can recognize from data, so it performs best when exposed to a larger sample. It also becomes better with more data exposure. A subvariant of ML is called Deep Learning (DL). DL uses convolutional neural networks (CNN), which correspond to algorithms based on brain neurons consisting of several hidden layers (multilayers). These CNN algorithms allow the determination of relations between given data (the input) and the results (the output). In this way, DL plays a key role in recognizing patterns drawn from large amounts of data.[15] The relationships between these various concepts are schematized in the following figure (figure 1).

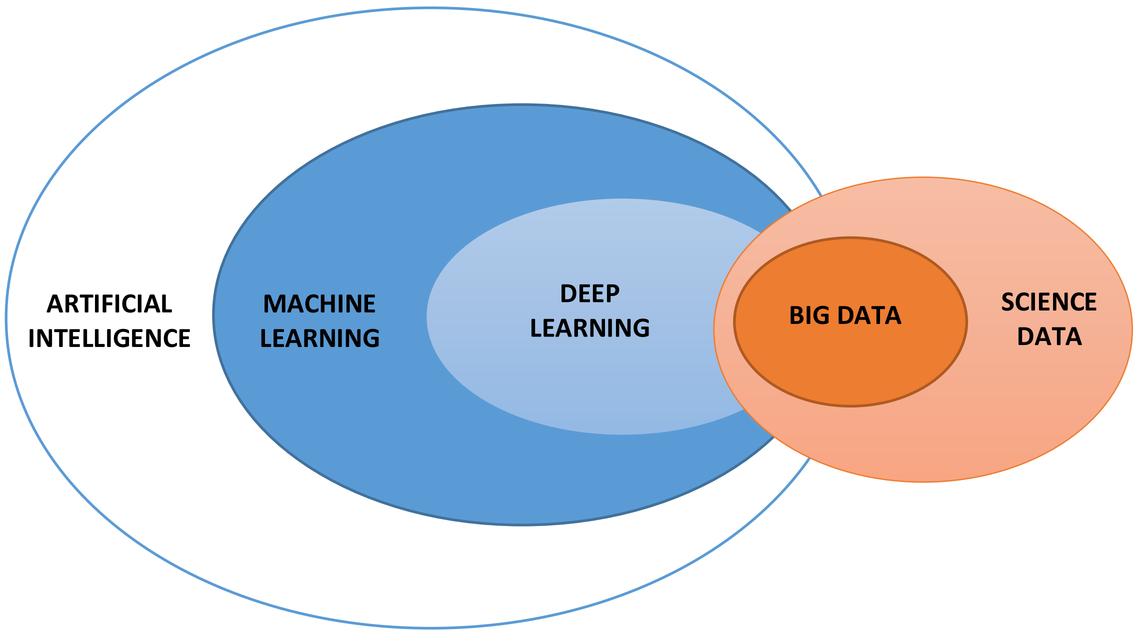

Figure 1: Artificial intelligence components.
(Based on *Drukker et al.*)[15]

Arthur L. Samuel was an electrical engineer and a pioneer in the use of AI, namely in the application of ML models. He studied the application of these models to checkers game, in an attempt to teach the computer how to detect the best moves in a short time and win the best play of the season, on a paper published in 1959.[18] The aim of ML in the field of Medicine is to apply it to tasks performed by healthcare professionals in order to help improve workflow, minimize error, improve results and take less time while performing them, particularly those which require repetition and that are more time consuming. ML is currently the main AI technology used to develop models to aid in obstetric US.[1] At Oxford University, one of the first obstetric models using ML was developed, the ScanNav™ assist, which is currently used in clinical practice. This model evaluates each image obtained during the second-trimester US and ensures that all structures have been evaluated and that the correct plane was obtained during this evaluation, all according to the ISUOG guidelines.[19]

1.4 Machine Learning Algorithms

ML algorithms are divided into three major types. **Supervised learning algorithms**, in which the learning is carried out under the supervision of a human, in the form of labelled objects. **Unsupervised learning algorithms**, on the other hand, learn without human interference or labelling. There is yet another type, **reinforcement learning**, which involves initially using labelled data and later inputting unlabelled data.[20,21]

Nowadays, the most commonly used algorithms in medical imaging are supervised learning algorithms, since they are generally best used when there is less amount of data, and when the input data is clearly labeled.[20] The main examples of supervised machine learning are listed in the table 1 and described briefly in the following text.

Table 1: AI supervised learning algorithms.

Supervised learning algorithms	Support vector machine
	Naïve Bayes
	Logistic regression
	K-nearest neighbors
	Neural networks

(Based on Erickson *et al.)*[21]

Support vector machine (figure 2) can divide input data into different classes with the greatest possible distance between them. This algorithm can do this splitting of data using a non-linear function, categorizing data that is not linearly distinct.[21] Thus, the error associated with the classification is smaller when there is a greater distance between the classes.[20] In obstetric US it can aid in facial recognition algorithms. It can also be used in handwriting detection and fraud associated with bank cards.[22]

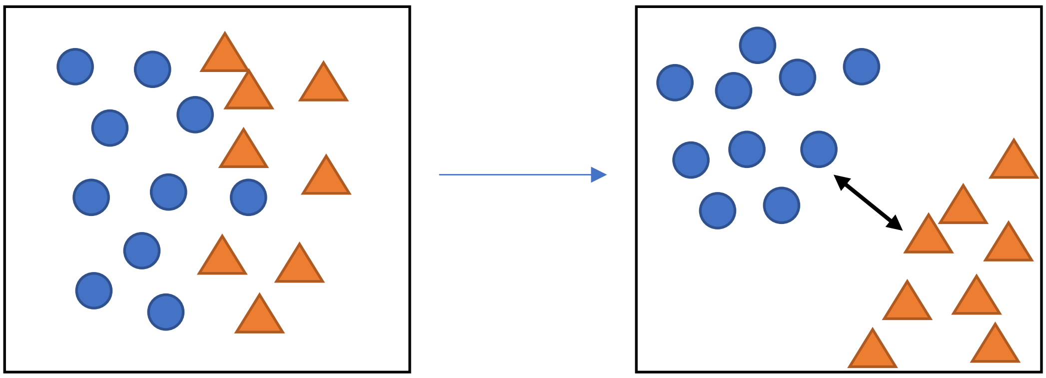

Figure 2: Support vector machine algorithm.
(Based on *Mahesh et al.*)[20]

Bayes' theorem suggests that the probability of a given event occurring comes from a series of events that are related to each other. This is one of the oldest ML algorithms.[21] The **Naïve Bayes algorithm** is used for classification and implies independence between the features present in each class. Thus, the organization of the input data into clusters/classes depends on the conditional probability of forming those same classes.[20] This algorithm can be used especially when the amount of data provided is small, allowing even more robust results.[21] However, in the real world, features are often associated with each other, so this can cause errors.[21] This algorithm can be used in models that make recommendations and in the detection of disease recurrence and progression after therapy.[22]

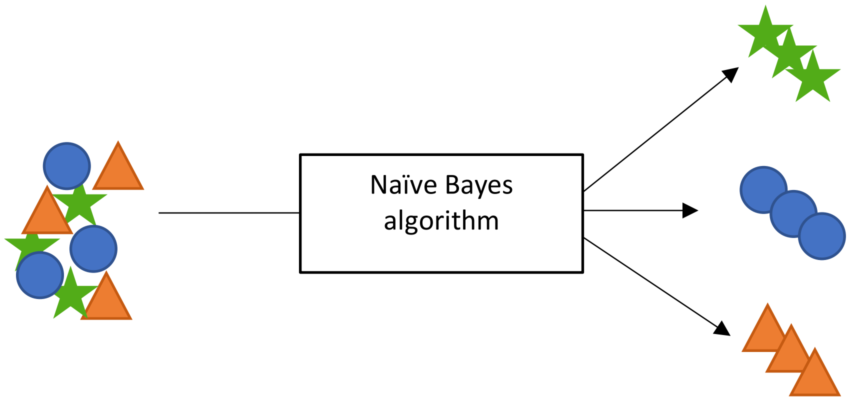

Figure 3: Naïve Bayes algorithm.
(Based on Erickson *et al.)*[21]

Logistic regression allows, based on input data features, to classify outcomes in a binominal way, i.e., through the probability of an event occurring (1) or not occurring (0). It has the advantage of being efficient and easy to implement. In addition, artifacts in data usually do not affect its performance.[22] It can be used in predicting the risk of developing a certain pathology, suggesting a diagnosis, and knowing the associated mortality.[22]

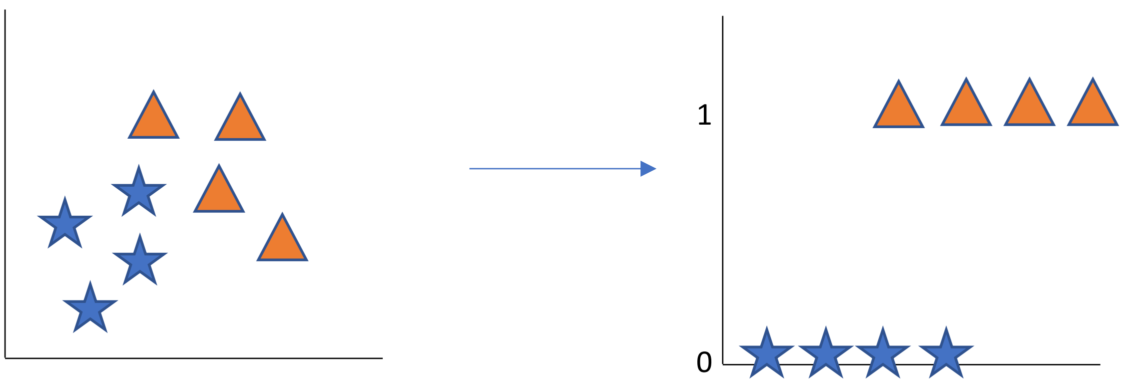

Figure 4: Logistic regression algorithm.
(Based on *Ray et al.*)[22]

K-nearest neighbors' algorithm: When an unknown object is introduced on the model for classification, an input vector (set of characteristics attributable to that object) is classified. The similarity of the characteristics is what will allow the classification, that is, the proximity of the known objects to the unknown one will determine the class to which it belongs, always depending on the number of neighbors to consider (k).[21] This algorithm can be used to do video or image recognition, as well as to develop systems that make recommendations.[22]

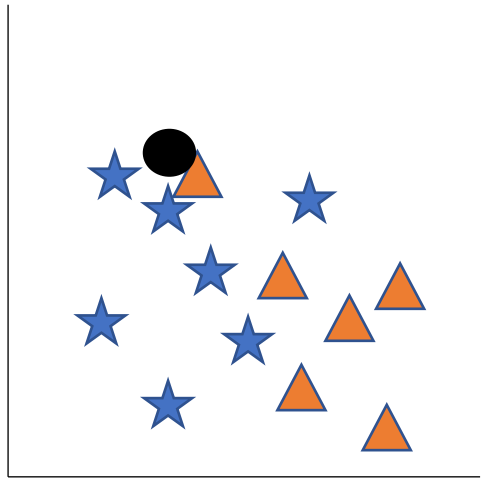

Figure 5: K-nearest neighbors' algorithm.
(Based on Erickson *et al.*)[21]

The **neural network algorithm**, as mentioned earlier, allows a relationship to be established between the input data and the obtained results (output). The difference between the obtained and expected results is used to update the data, which in turn will be improved so that this difference is smaller. This process is carried out until the error reduction is no longer significant. [21]

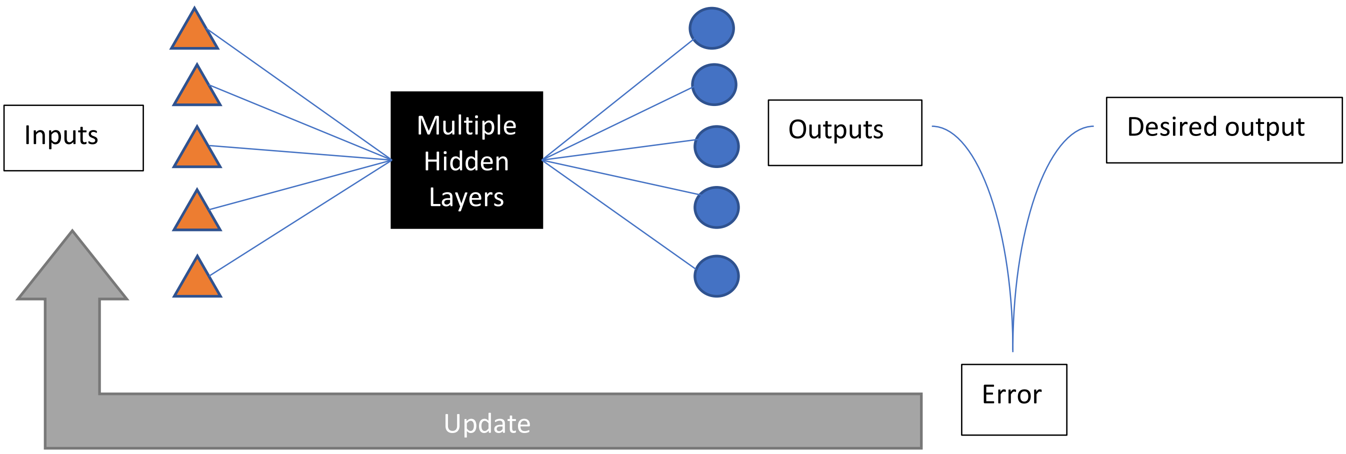

Figure 6: Neural Network algorithm.
(Based on Erickson *et al.*)[21]

Thus, with the creation of these algorithms, ML will allow automatic evaluation of new data. In the specific case of obstetric US, automatic classification of images, characterization of whether they belong to the normal spectrum or whether they detect a deviation from normality, and measurement performance could be automated. In this way, all the manual repetitive tasks performed in a fetal US scan could be avoided. Also, less exhaustive training would be an option in places with less resources, and telemedicine could be implemented in regions with less access to specialists in Obstetrics. This would also allow increased diagnostic capability, error reduction, easier workflow for clinicians and, overall, better health outcomes.[11]

To summarize, we aim to investigate on the impact of AI models application on mid-trimester US, in particular on the ability to diagnose fetal malformations and deviations from normality. We will also analyse AI´s ability to obtain automatic planes and analysis of fetal biometry (for fetal weight and GA estimation), fetal well-being, placental abnormalities, US workflow and the possibility of shortening the exam duration.

2. Materials and Methods

For literature search, three databases (Pubmed, Elsevier and WebOfScience) were accessed in December 2022.The equation used was: ("ultrasound"[mesh]) AND "artificial intelligence"[mesh]). A total of 4265 results were obtained and after removing duplicates, 2940 articles remained. Articles published between 2012 and 2022, in English, Portuguese or Spanish, that related the application of AI models in second trimester obstetric US were defined as inclusion criteria. Meta-analyses, systematic reviews, randomized clinical trials, narrative reviews, cohort studies, case-control studies, and cross-sectional studies were included. Redundant articles, study protocols, commentaries, and articles without available full versions were excluded.

After applying these criteria, the selection started with the evaluation based on title and abstract that allowed selecting 196 articles. Then, based on the full text, 160 articles were excluded. After checking the references of the selected articles, it was possible to add 6 more articles. In addition, 11 articles on guidelines for performing obstetric US and data about fetal US diagnosis, and 7 articles on understanding and defining AI and the associated concepts were included. Thus, 63 articles were used in this narrative review (Figure 7).

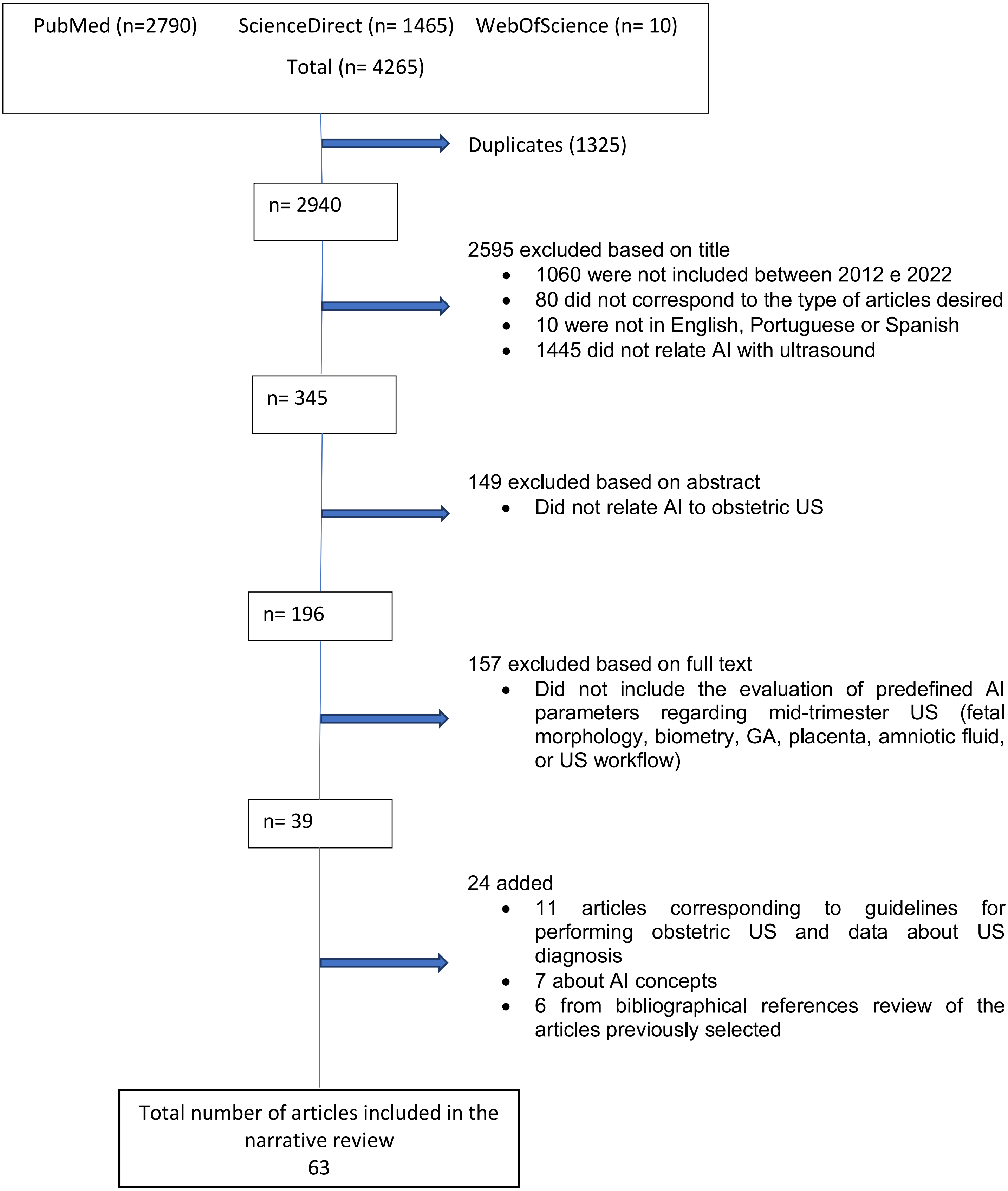

Figure 7: PRISMA flow diagram of materials and methods.

3. Discussion

3.1 Fetal Anatomy

The fetal anatomy evaluation is one of the most important components of the second trimester US and it is used to identify the existence of anatomical defects. This allows couple counselling and, if adequate, possible in-utero or neonatal interventions that might improve outcomes.[3]

3.1.1 Heart

Congenital heart defects (CHD) are the most common fetal anatomy abnormality. Although it only affects 1% of all births, about 20% die within the first year of life due to CHD.[23] It is also important to highlight that the majority of CHD are detected in low-risk pregnancies.[24]

The reasoning to create computer automatic models using AI to recognize CHD is due to its low diagnostic sensitivity (30%) and specificity (40-50%) when US is performed by expert sonographers. Some factors contributing to these outcomes are fetal heart movement, it´s small size, and variation of the quality of the obtained images, whether by the same or different observers.[1,23] ISUOG guidelines suggest the evaluation of five different views of the heart to safely enable diagnosis of CHD. These include the four-chamber view, three-vessel view (3V), three-vessel trachea view (3VT), left and right ventricular outflow tracts and abdominal situs. All these planes would theoretically allow the diagnosis of about 90% of CHD cases. In practice, however, lower sensitivity and specificity are observed, as mentioned above.[3] The main CHDs that can be diagnosed in mid-trimester US are aortic coarctation, hypoplastic left heart syndrome, tetralogy of Fallot, transposition of great arteries, double outlet right ventricle and truncus arteriosus. It is also important to detect defects in the ventricular septum, which can be difficult to do because they can sometimes be very small (1-2 millimetres).[25]

As mentioned, there are five heart US planes which are gold standard for heart visualization and CHD diagnosis. *Nurmaini et al.* have proposed a method for automatically segmenting these standard views and to identify heart anomalies at the same time (atrial, ventricular and atrioventricular septal defects). They concluded that they were able to get the model's detection of CHD to be like that of the clinicians', but their precision was higher on images belonging to the same fetus and somewhat lower when comparing images from different fetuses (mean average precision of 98,3% vs 82,42%). According to the authors, this could be due to various factors, such as image quality, wide variability between hearts from different fetus, and defect complexity (particularly regarding it´s extension).[23]

Arnaout et al. developed an ensemble of neural networks (figure 6) which aimed to identify the recommended cardiac views. This model´s goal was also to assist on identifying normal hearts or CHD. To do this, they used several images sets with different proportions of representative CHD images, including one set with a number of images of fetuses with CHD equivalent to that observed in the general population (about 1%). Tetralogy of Fallot and hypoplastic left heart syndrome were the main CHDs present in the study database. They concluded that for the former, the best planes for diagnosis were 3V and 3VT, while in the latter, anomalies were observed in all planes. For both, there was an increase in diagnostic sensitivity relative to that noted on current clinical practice (71% vs. 50% and 89% vs. 30%, respectively). Globally, the model achieved an area under the curve (AUC) of 0,99 (meaning that the precision obtained was almost perfect). The ability of the model to detect CHD in fetuses that actually had heart anomalies was similar to that of physicians, and lower quality images could be used for this while maintaining diagnostic quality when the model was used outside the hospital setting.[3]

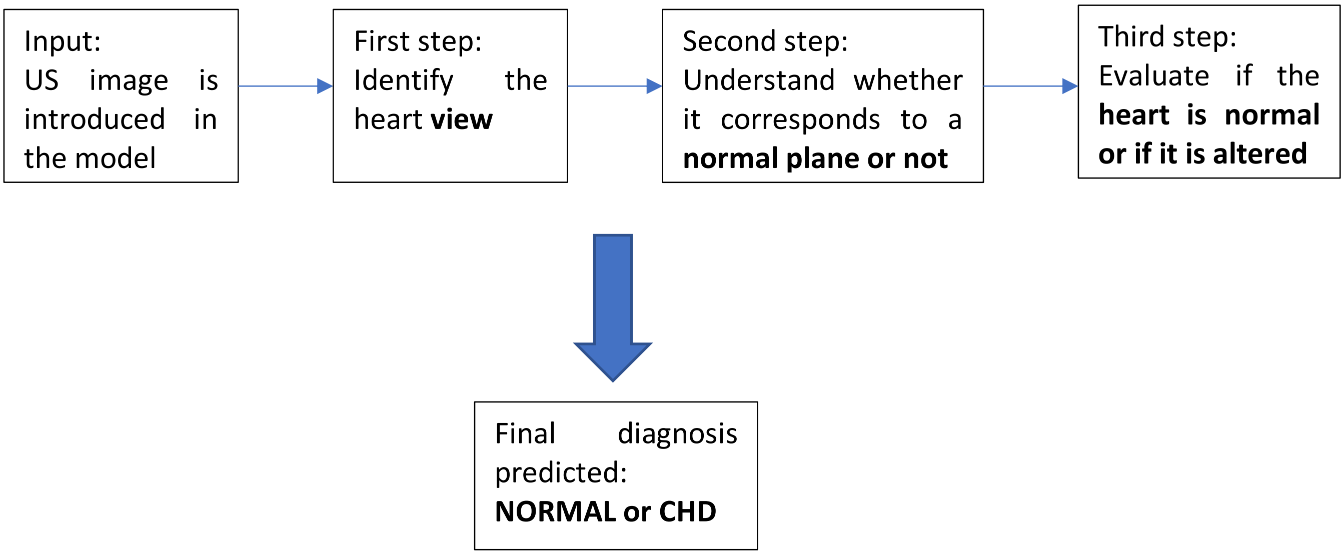

Figure 8: Steps of the ensemble neural network.
(Based on *Arnout et al.*)[3]

On the same cardiac segmentation model to detect CHDs, an automatic segmentation model to obtain fetal biometrics, namely cardiothoracic ratio, cardiac axis, and fractional area change applied to each of the four chambers, was created. According to the author, the need to obtain a model with this function arises from the fact that these measures are not yet benchmarked. The objective was to understand how these measurements vary in normal hearts and hearts with tetralogy of Fallot and hypoplastic left heart syndrome. With statistically significant results, they noticed that the cardiac axis was increased when one of the CHDs was present (45 degrees in normal hearts and 63 degrees in tetralogy of Fallot).[3]

As mentioned before, ventricular septal defects are the most frequently detected CHD. *Dozen et al.*, through image segmentation, improved a previously developed automatic model that specifically recognizes the ventricular septum. This model allowed better similarity values (mean Dice coefficient) when compared to previous existing models. Although the latter was only tested on normal ventricular septa, the results are promising for achieving better CHD diagnosis in the future.[26]

Yeo et al. evaluated the possibility of using an automatic model they had previously developed, which generated standard visualizations of the heart, in identifying four CHD: aortic coarctation, Fallot tetralogy, great vessels transposition, and pulmonary atresia with intact ventricular septum. Using this model on US cardiac images picturing these abnormalities, they concluded that a change in the normal anatomy of the heart was detected in all of them.[24]

3.1.2 Thorax and lungs

In the second trimester US, thorax and lungs should be evaluated regarding their shape and transition into the abdomen. In addition, all ribs, lung morphology and, if possible, the diaphragm should be assessed. The shape of the chest wall can reflect certain pathological situations. Its deformation can be related to the presence of pectus excavatum, while if it has small dimensions, it can be related to bone dysplasia conditions. In addition, it´s evaluation can also aid in the detection of CHDs and diaphragmatic hernia when assessing the cardiac axis.[6]

Shozu et al. developed a deep learning AI model to automatically perform chest wall segmentation using the 4 chamber-view plane and thus evaluate the relationship between the structures that are part of the thorax, their locations, and relative positions. These factors, their deviation from normal, and their automatic detection, allow to predict the possibility of a congenital anomaly. Furthermore, a correct segmentation of the chest wall allows the construction of new automatic models in which it is possible to obtain essential planes in the evaluation of the existence of congenital anomalies. To do this, they used a method that combined serial information from US videos with the shape of the chest wall. They compared this new model alone with the association between the one developed and two others that had already been developed for image segmentation. They were able to get a higher Dice coefficient with both the associations, 0.654±0.005 (compared to 0.610±0.016 from another model isolated) and 0.633±0.004 (compared to 0.582±0.007 from another model isolated). The authors also tested the application of existing segmentation models to the fetal lung. The Dice coefficient obtained was greater than 80% for both models.[27]

3.1.3 Central Nervous System

Central nervous system anomalies have an incidence of about 2 cases per 1000 births. Therefore, brain evaluation and its possible structural abnormalities is mandatory in obstetric US. According to ISUOG guidelines, there are structures that require an obligatory evaluation, such as the lateral ventricles, cavum septi pellucidi, thalami, cerebellum and cisterna magna.[6] There are some limitations regarding fetal brain US evaluation, such us the shadow produced by the skull, anatomical physiological changes of the brain throughout pregnancy, and sometimes, the difficulty of correctly positioning the probe according to fetal orientation.[28] Given the disadvantages mentioned above, the CNS evaluation starts as early as the first trimester, where various CNS anomalies can be detected, such as neural tube defects, anencephaly, encephalocele and spina bifida. Malformations of the posterior fossa are another type of anomaly diagnosed in the first trimester. Later in the second trimester, other anomalies can be detected, including ventriculomegaly and midline anomalies, such as holoprosencephaly, hydrocephaly, septo-optic dysplasia, and corpus callosum agenesis.[29,30] When these abnormalities are detected, more imaging scans may be necessary, such as neurosonography or brain magnetic resonance imaging.[30]

Xie et al. developed a DL model that aimed to distinguish between normal and abnormal US fetal brain images. The results showed an increased potential to classify images, with a 96% accuracy rate. The detection of images picturing brain anomalies was 96.9%, with a specificity of 95.9% (AUC of 0.99). It is also important to highlight that the time required for automated abnormality annotation (through segmentation, classification, and localization) was about 1,08 seconds.[31]

Regarding 3D fetal brain US, *Huang et al.* proposed an automatic model using VP-Nets (view-based projects networks), a three-view based CNN, for detecting five fetal brain structures (cavum septi pellucidi, thalami, lateral ventricles, cerebellum and cisterna magna) in 3D neurosonography images. The model achieved an accuracy over 60% for detecting the referred structures. The lateral ventricles were the smaller structures to be identified, and that is also the reason they were the ones with the poorest detection accuracy.[28]

On the other hand, the model proposed by *Montero et al.* aims to generate new artificial images (data augmentation) to improve US brain image classification, hence removing the error associated with manual interaction with the images. The main goal was to distinguish between the transthalamic plane and the transventricular plane. They used generative adversarial networks (GAN) (figure 7) that, by using both real images and artificial images, could classify fetal brain images with a greater accuracy and distinguish between both planes.[32]

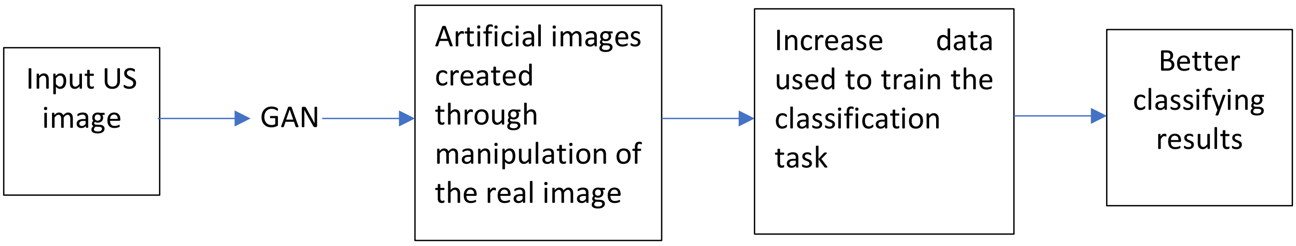

Figure 9: Generative adversarial network.
(Based on *Montero et al.*)[32]

Hesse et al. developed a model based on a CNN algorithm for automating the segmentation of subcortical structures in 3D US images. The structures specified in the article were the choroid plexus, horns of the posterior lateral ventricle, cavum septum pellucidum and cerebellum. Through this technique, the goal was to evaluate the adequate development of these structures along fetal growth, so that possible neurological dysfunctions could be detected. With this work they were able to define the expected growth curves for each subcortical structure obtained in the mid-trimester US, and with that identify appropriate brain development with increasing GA. They concluded that, although the choroid plexus undergoes a small variation in absolute size during the second trimester, it undergoes a large variation in size relative to the whole brain from being 12% of the total brain volume to 3% in the third trimester. On the other hand, the cerebellum, despite experiencing a large increase in the second trimester, has a less marked volume variability, representing 1.5% of the total brain volume in mid-trimester, and 2.5% in the following trimester.[33] In fact, the cerebellum is an important structure to be assessed during fetal brain visualization according to ISUOG guidelines on sonographic examination of the CNS.[30] The evaluation of this structure is particularly crucial in the mid-trimester US scan, since the skull has not yet reached maximum calcification, and it can more easily be observed. *Shu et al.* developed an automatic segmentation model of this specific structure using a neural network algorithm, based on U-net, using mid-trimester fetal cerebellum images. The importance of its segmentation is because it is a structure that undergoes various structural changes and reaches an irregular morphology when it achieves maturity. To see how close the automatic segmentation was to the ground truth, they used two similarity indices to evaluate the model (Jaccard score and Dice similarity coefficient) obtaining values of about 86% and 91%, respectively. When comparing with other image analysis automatic models, this one allowed improving both values.[4]

3.1.4 Face

According to ISUOG guidelines regarding fetal face evaluation, the mandatory parameter to be assessed is the integrity of fetal lips and nose. That is because cleft lip and palate are one of the most frequent facial abnormalities. In addition, the orbits, nostrils, and the median profile of the face should be observed whenever accessible. There are three crucial US planes used for fetal face examination. By being able to evaluate the distance between the two eyes, the ocular axial plane allows the detection of hypertelorism, distance between both eyes greater than the 95th percentile, or hypotelorism, the same distance less than the 5th percentile. Both can reflect various pathologies, such as chromosomal alterations or genetic syndromes, and it is also possible to detect ethmoidal encephalocele and facial haemangioma (hypertelorism) and holoprosencephaly and microcephaly (hypotelorism).[34] The median sagittal plane allows cleft lip detection (that can also be done by the nasolabial coronal plane), frontal bossing, micrognathia and nasal bone anomalies.[6] Each plane requires a certain probe positioning for accurate observation. For example, both crystalline lens and both eyeballs should appear with the same size in the axial plane.[35]

Wang et al. developed an automated model using a texture fusion method for fetal facial characteristics recognition and classification. They concluded that the model met the objective with quality and high effectiveness, achieving an accuracy and precision values greater than 94%.[35] *Lei et al.* also developed an automatic plane recognition model of the fetal face using Fisher vector to identify the above-mentioned US planes (axial, coronal and sagittal). They were able to recognize standard planes with an accuracy of 99% and an acuity of 93%.[36]

Also, with the aim to improve adequate acquisition and evaluation of standard planes of the fetal face, *Yu et al.* developed an automatic DL model using CNN to achieve a thorough evaluation of the US image. This new method obtained true positive values greater than 98% and accuracy values greater than 99% for the axial and coronal plane. When compared to the previously described models, it obtained the highest values of accuracy, true positives and precision (greater than 96%).[37]

For craniofacial assessment there are several measurements that can be used and their variation from normal can be discrete and difficult to observe in US images. *Tsai et al.* developed an automatic image segmentation model that tested the following measurements: BPD, occipitofrontal diameter (OFD), interorbital diameter (IOD), bilateral orbital diameter (BOD) and the distance between vertex and nasion (VD), using 3D US scans. When compared to manual measurements taken by an expert, the correlation value was higher than 0.8 in all, reaching a maximum value in BPD (0.99). This showed that the automatic model is very close to the manual one, and that the model took only 30 seconds to obtain the five measurements.

The face contours obtained automatically and manually by image segmentation differed by only 0.6 mm. It should be noted that the study was conducted on a very small number of healthy pregnant women (n=11). In addition, the limitations of the model related to fetal movement during the time of US and the difficulty it may be to apply it in later stages of pregnancy (due to difficulty in obtaining an image with the entire fetal head) were defined.[38]

3.1.5 Abdomen

The fetal abdomen also requires a detailed evaluation because of the large number of structures that must be assessed. In addition to assessing the stomach, its integrity and location, the insertion of the umbilical cord should also be assessed. In a systematic evaluation the stomach (left side) should be visualized, then the umbilical vein, and further to the right the gallbladder.[6]

A model was developed to automatically detect AC planes using ML. The aim was for the model to detect the stomach and umbilical veins on 3D US. Compared to the manually analysed planes, the model was able to achieve a sensitivity of about 91.3%. In addition, they were able to obtain a positive predictive value of just over 76%.[39] *Ahmed et al.* developed an automatic model for detecting abdominal planes in 3D US images. This model was built on the basis of the eye movements of the sonographers, in particular the evaluation of the planes that were fixed by their gaze and possibly less important planes that did not catch their gaze. The main objective was the detection of standard planes containing the abdominal wall, the stomach and the umbilical vein. They were able to obtain an accuracy value, i.e., correctly detected structures, of 92.5%. In addition, the software took less than 11 seconds in evaluating each abdominal US image.[40]

For standard plane detection of AC, *Cai et al.* developed a CNN-based model. They also used the sonographer's gaze and its evolution throughout the examination, which points of the US scan were fixed and which ones were uninteresting. Thus, they concluded that the association of the sonographer's gaze allows the increase of the model's performance in detecting the planes containing the AC, with a precision rising from 73.9% to 96.5%.[41]

3.1.6 Genitalia

In the mid-trimester US, the evaluation of the fetal genitals should be done to assess their normality. Gender identification can be done according to the parents' wishes and local regulations.[6]

Using neural networks, *Maysanjaya et al.* developed an automatic model to classify fetal sex. To do this, the images that made up the input data were grouped according to classes (clusters). This was achieved by analysing the characteristics of each image and assigning them for each cluster. Supervised and unsupervised algorithms were used. The ability to correctly classify the images accounted for only about 61%. The identification of images of male fetuses was more accurate when compared to the identification of female fetuses.[42]

Tang et al. created an automatic model to classify the fetal gender using points of interest identified through features of genital organs, such as the saliency. This was important due to ethical issues in some countries related to determining the sex of fetuses and sex-selection abortion. They were able to achieve true positive values of about 80% and true negatives of 83% in identifying the fetal sex. Also, each image required about 0.45 seconds to be evaluated and provide a result.[43]

Another method of automating the classification of fetal sex was developed by *Kaplan et al.* Initially, the features considered more valuable were automatedly selected from US images to be classified. Five different ML algorithms were used for this classification. They obtained high accuracy when using k-nearest neighbors and support vector machine, respectively 99.1% and 98.5% with AUC greater than 0.88.[44]

3.2 Fetal Biometry

Fetal biometry in mid-trimester US consists of measuring the BPD, HC, AC and FL. These are the measurements used to make estimates of fetal weight (EFW). Hadlock's formula, which uses these measurements, is currently the preferred method for calculating EFW and the one that gives better results. Thus, it is possible to diagnose whether a fetus is small for gestational age, normal, or large for gestational age, depending on the percentile in which it´s EFW is included. Fetal growth may be compromised by the presence of maternal risk factors (the most frequent being uteroplacental dysfunction) and by possible fetal conditions such as infection, aneuploidy, and genetic diseases. The existence of growth pathology, namely fetal growth restriction, represents a high risk for the fetus and may be the cause of about 30% of stillbirths. Therefore, its early diagnosis during pregnancy, namely during the second trimester, allow for therapeutic and surveillance measures to be undertaken in order to prevent its harmful effects.[6,45]

Fetal biometry can also be used for calculating the GA when a crown-rump length was not previously obtained.[6] Pursuing better GA accuracy to optimize prenatal care interventions dependent on GA, such as fetal lung maturation using corticosteroids or labor induction, is a relevant issue in Obstetrics. Furthermore, pregnant women who live in countries with limited access to health care usually resort to obstetric care later in pregnancy.[13] This is aggravated in the context of high-risk pregnancies.[46] As pregnancy progresses and the fetus grows, fetal biometry has a wider variation, leading to less accurate GA predictions.[47]

In order to identify standard planes for biometry assessment, *Cai et al.* developed an automatic model by using the gaze of sonographers. The planes to be detected were the ones that allowed measurements of HC, AC and FL. This model was able to obtain an accuracy (F1 score) values, for each of the biometric planes, greater than 81%.[48]

Smartplanes® is an AI model developed to identify image planes on US. *Grandjean et al.* tested the feasibility and reproducibility of this software in obtaining BPD and HC measurements in 3D-US, specifically by identifying the head transthalamic plane. They concluded that the reliability for the measurements obtained by different observers was high, and that the values obtained by the automatic method were strongly comparable with those obtained by experienced sonographers. The process of capturing the images and making the automatic calculation took less than 10 seconds.[46]

Burgos-Artizzu et al. proposed an AI model for automatic GA estimation by evaluating brain texture and size observed in US. They analysed the AI model alone and the same model associated to fetal biometry evaluation. In second-trimester US, the combination between the first AI model and fetal biometry (BPD, HC, AC and FL) showed the lowest average absolute error (2,44 days) and a 95% confidence interval error of 6,7 days (vs 7 days when using only fetal biometry parameters). The difference between the two was even more significant for small for GA fetuses (14.8 vs 18.5 days).[49]

By identifying and segmenting the skull, abdomen, and femur, *Dan et al.*, developed an automated model to estimate GA using DL. The model consisted of a first stage of segmentation of the areas mentioned to obtain the biometric measurements. Then a regression was used to estimate the GA. The mean absolute error, which is the mean difference between the GA values estimated by the model and the ground truth, was lower than 1 week using this model, but with a high standard deviation (0.57 weeks). Nevertheless, this value was close to the experts, which made their predictions with an even higher standard deviation (1.28 to 1.35 weeks). Furthermore, each individual prediction took 0.37 seconds.[50]

Yang et al. proposed, using an end-to-end CNN, an automatic model to segment the contour of the fetal skull and calculate HC. They achieved very similar and comparable to the state-of-the-art HC measurements. The mean difference value obtained was 0.08 ± 2.37 milimeters.[51]
Throughout fetal development, the brain changes from being a smooth surface (in the second trimester) to having texture, created by the gyri on the third trimester, when it is already similar to the adult brain. Knowing that assessment of fetal brain patterns may allow a correct prediction of GA, *Namburete et al.* proposed an automatic model based on ML. This semi-automatic model was used to divide the fetal brain into various anatomical parts, which were

manually indicated by experts. Then, the main characteristics of each fetal brain anatomical part according to GA were detected and used in the learning process. This model presented a confidence interval of 1.68 days GA prediction on the second trimester. The automatic method outperformed the clinical method, by obtaining the average of three HC measurements, by 0.91 days. From the data, we can see that the associated errors are lower in the second than in the third trimester. They identified several brain regions as extremely useful for more specific GA estimation, such as the sylvian fissure, cingulate sulcus and callosal sulcus.[13]

Wang et al. developed a new model using CNN with the goal of measuring HC. They intended to eliminate the blunts from the image, namely the skull bone. In this way, they were able to obtain a similarity coefficient of 98.2%, with a *Hausdorff* distance value of about 1.22 milimeters.[52]

An important parameter used for biometry is the FL. However, there are some limitations on obtaining a correct segmentation of the femur image. The automatic models developed so far were only able to use the most distal points of the femoral bone, whereas on the clinical set, FL is obtained by excluding the femoral epiphyses.[53] *Zhu et al.* proposed to compare two automatic models that aimed to automatically detect femoral endpoints and measure FL. In addition, they compared the two models with measurements obtained manually by sonographers. One of the models was based on a random forest regression (a ML method) and the other was based on SegNet (a DL method). When comparing each model with manually obtained measures, the difference between FL obtained with random forest regression method was 1.23 ± 4.66 millimetres and with the SegNet method was 0.46 ± 2.82 millimetres. They concluded that the DL model achieved better results compared to the ML model.[53]

3.3 Placenta

Mid trimester evaluation of the placenta is very important, namely assessing its location in relation to the internal cervical os, and appearance. Some anatomical normality deviations that can be found are haemorrhage, anechoic cysts and masses, such as chorioangioma.[6]

Hu et al. developed an automated CNN-based placenta segmentation model, named U-net. Layers of the CNN were replaced by ones that allow the identification of artefacts in US. These are mainly acoustic shadows, which do not allow clear observation and analysis of the images. Approximately 18% of the images in use had this artefact. The aim was to evaluate the model's ability to identify the placenta when it was not continuous in the image (i.e., interrupted by acoustic shadows). They were able to obtain a Dice coefficient (comparing the automatic segmentation with the manual segmentation performed by a specialist) of around 92%, the highest compared to other models without artefact removal. Some errors occurred due to the similarity between structures that can resemble the placenta, such as the myometrium.[54]

Lei et al., to overcome the subjectivity inherent to US evaluation of the placenta, created an automated method for assessing placental maturity using multi-layer Fisher vector. This allowed an entire automatic placental assessment, not layer by layer. Through the evaluation of placental maturity, it was possible to infer about its functional capacity, and consequently about fetal health, growth and development. With this model they were able to stage placental maturity with a 98% sensitivity, a 94% specificity and an accuracy of 96.8%.[14] A model developed by *Li et al.* using a dense descriptor with the four standard stages of placental maturity, with the same objective as above, achieved a 99% sensitivity and a 87% specificity, with an accuracy of about 87%.[55]

Qi et al. developed a model to identify and localize the placenta, myometrium and subcutaneous tissue in US images. Therefore, by observing the relationship between these structures, a diagnosis of abnormally invasive placenta may or may not be made. These are structures that can be difficult to differentiate and are subjectively assessed by the clinician, hence the importance of automating this process. The authors created a CNN-based model for the purpose of describing these anatomical structures and the various interactions between them. They were able to obtain a top-1 error rate of less than 0.087, which corresponds to the proportion of times the model incorrectly classified the three structures mentioned before. Thus, they concluded that the model could locate all the structures correctly.[56]

The abnormally invasive placenta usually runs its course with characteristic lesions - the lacunae - that can be seen on the US scan. This abnormality leads to qualitative image changes, namely transparency on US, and quantitative changes, such as the number of lesions and their sizes. It is necessary to pay attention to whether these changes are pathological, that is, if there are several lesions, large and with an irregular contour. The same authors developed an automatic model for detecting this change using a layer aggregation method based on deep aggregation. They obtained a mean average precision value of 35.7%, which was the highest value when comparing to other models.[57]

Recalling the difficulty of having clinicians specializing in US in developing countries, *Schilpzand et al.* developed an automatic model that could be used by connecting a smartphone to low-quality US equipment. The main purpose of the model was to detect the existence of low-lying placenta or placenta previa. Many images did not contain the cervix so that it could be used as a reference to distinguish between the two pathologies, but the aim was to indicate whether there was a risk of matching one of the situations. To do this, a DL model first performed segmentation of the placenta, followed by image classification according to whether it was considered normal or abnormally positioned. It was possible to obtain a Dice coefficient for segmentation of 0.84, and a sensitivity and specificity of over 80% was obtained for classification. Furthermore, the real time for image segmentation varied between 17 and 21 miliseconds.[58]

3.4 Amniotic fluid

Amniotic fluid quantity is used to assess fetal development and well-being. According to international guidelines, it can be assessed subjectively; objectively, by measuring the depth of the largest fluid pocket; or by using the amniotic fluid index (AFI), which combines the measurements from the amniotic fluid pockets regarding the four quadrants of the pregnant uterus.[6] Segmentation of the amniotic fluid pockets in US scans is essential, as it allows their boundaries to be identified, making it possible to determine their dimensions more objectively. Its manual performance by a sonographer depends on several factors, such as amorphous features and artifacts inherent to the US technique.[59]

Cho et al. developed a DL automatic method to segment the fluid pouches in US images and, with those results, calculate the AFI (AF-net). They calculated the Dice similarity coefficient obtained by the automatic model and the one obtained by a sonographer, which was 87%. The obtained AFI measurements by the AF-net model showed a mean absolute error of 2.67 ± 2.99 millimetres and a mean relative error of 0.018 ± 0.023 millimetres, which were the lowest error values obtained when compared to two experienced sonographers.[59]

To improve the previously described model, *Sun et al.* developed an automatic model capable of recognizing and segmenting image interferers, namely reverberation artifacts. To do this they used three different databases, one containing only images of AF pockets with artifacts, one without, and one containing both types. They obtained a model with a better Dice similarity coefficient (0.8599±0.1074) compared to AF-net alone (0.8557±0.1072) and with significantly greater sensitivity (0.8073 vs 0.8180) for all data (with and without the artifacts). However, the Dice coefficient obtained with this new model was higher when it was applied in the dataset containing the AF pocket images with artefacts (0.9042±0.0361) compared to the datasets containing images without artefacts (0.8510±0.1165) and to all data (0.8599±0.1074).[60]

3.5 Ultrasound workflow analysis

Performing a second trimester obstetric US requires careful evaluation of several structures. Although there are now several guidelines to make it less variable between hospitals and clinicians, the time and order to perform this exam is not established.[61] Also, the quality control of images obtained in fetal US, which is a time consuming and labor-intensive task, has to be performed manually by very experienced sonographers. This leads to fatigue and a greater likelihood of mistakes.[5] Therefore, and in addition to the direct application of AI in US, algorithms can also be used to evaluate the performance of sonographers, how to improve it, as well as to ensure the best possible quality of images, relieving their workload.[61]

Sharma et al., through a retrospective analysis, evaluated the workflow of full-length US performed in the second trimester. They used a database containing about 200 hours of videos from second trimester exams to temporally segment images and identify what is valuable in each one using a DL method. Other goal was to develop an automated model using CNN to assist in performing US (providing feedback, suggesting new ways to improve image acquisition, and reducing the time to perform US). The model obtained a strong correlation between manually analysed images and those identified by the machine. For the detection of meaningful images, the model achieved an accuracy of 91.7% in the statistically significant cross-validation and an accuracy of 76.4% in the manual retrospective evaluation made by 3 different sonographers.[2]

Also, with the goal of understanding the workflow during second trimester US, *Drukker et al.* developed an automated DL neural network model using prospective observational data. By understanding and evaluating the way the specialists perform this exam, it may be possible to create models that assist in the interpretation of these scans. The main goal was to understand which structures were most often detected, which structure was chosen to start the scan, and which sequence was used. They concluded that the most frequently identified structure was the skull and brain, followed by the facial coronal plane and the abdomen. Most sonographers preferred to start the imaging exam by the skull and brain. They concluded that the order of imaging and assessment of structures was different for all sonographers, highlighting the enormous variability.[61]

Zhang et al., through an observational study, aimed to develop an automatic method to evaluate the quality of US images using CNN. They aimed to understand if the captured image planes were optimal for the correct evaluation of fetal structures. They concluded that, with this model, it was possible to evaluate and ensure the quality of images, requiring less than 1 second per frame. In addition, they concluded that this quality assessment was superimposable to that performed by the sonographers.[5]

To aid the interpretation of second trimester US, *Alsharid et al.* developed an automated model to describe the exam based on deep neural networks. To do this they used videos of USs, analysed the sonographers' gaze, and recorded their speech describing the images they saw. They concluded that these models, which integrate the different components of the sonographer (vision and speech) into their training, have created a model with the ability to provide more detailed and varied descriptions for scan videos of mid-trimester US.[62]

In order to understand the impact of AI on mid-trimester US in real-world clinical practice, a prospective study was conducted by Matthew *et al.* For this purpose, 23 healthy pregnant women were selected to undergo both AI and manual US scans over the course of one year. Plane detection, biometrics and automatically generated reports were provided by the AI model. They concluded that AI saved 7.62 minutes per scan. According to the sonographers, the use of AI helps them to perform the US scan by saving time, which they can use to look at the images more closely and to focus on the relationship between the obstetrician and the parents. Nevertheless, the difference in time taken to report the results between the two methods was approximately 0.10 minutes. The manual method took the least time (2.39 minutes). The automatic method reported 93% and the manual method 98% of the essential planes expected to be detected. As the number of planes obtained rose, only 73% were reported by the AI model. Regarding fetal biometry, only the HC measurement showed a statistically significant difference between the automatic and manual methods, with a mean standard deviation of - 2.44 millimetres and a mean difference from the GA forecast of 1 to 2 days.[63]

The following table (table 2) summarizes the possible applications of the AI models found in each component of the second trimester US scan. Table 3 corresponds to the observed uses of AI models in clinical practice.

Table 2: Summary of possible AI applications in several components of second trimester US.

Mid-trimester US components	**Possible AI application**
Heart	• Identify and classify CHD • Ventricular septum segmentation • Detection of abnormalities in images containing CHD
Thorax and Lungs	• Chest wall segmentation • Lung segmentation
Central Nervous System	• Identifying brain images with anomalies • Ability to generate artificial images • Establish the evolution/growth of brain structures with increasing GA • Structure segmentation i.e., the cerebellum
Face	• Fetal face standard planes recognition • Obtain craniofacial biometry
Abdomen	• Detection of structures as stomach and umbilical veins • AC segmentation using sonographers gaze associated with the automatic model
Genitalia	• Identify and classify fetal sex
Fetal Biometry	• Obtain HC, BPD, FL measures • GA prediction
Placenta	• Assess placental maturity • Segmenting and localizing the placenta • Detect existing lesions in abnormally invasive placenta • Removing artifacts from US images of placenta
Amniotic fluid	• Segmentation of amniotic fluid pouches • Removing artifacts from US images of amniotic fluid pouches
US workflow	• Compare sonographers workflow and automatic models • Analyse the structures and planes evaluated by sonographers • Evaluate and ensure image quality • Automatic description of a US scan

Table 3: AI applications in several components of second trimester US used in clinical practice.[64]

Mid-trimester US components	AI application	Technology name
Heart	• Automatic identification of fetal anatomy in standard US planes	SonoLyst™
Thorax and Lungs	• Automatic identification of fetal anatomy in standard US planes	SonoLyst™
Central Nervous System	• Correctly align and display recommended planes • Measurements for fetal brain assessment • Semi-automatic measurements of cisterna magna, posterior horn of the lateral ventricle and cerebellum	SonoCNS™ SonoBiometry™
Face	• Automatic identification of fetal anatomy in standard US planes • 3D skin lighting and shading techniques for enhanced image definition	SonoLyst™ HDlive™
Abdomen	• Automatic identification of fetal anatomy in standard US planes	SonoLyst™
Genitalia	• Automatic identification of fetal anatomy in standard US planes	SonoLyst™
Fetal Biometry	• Semi-automatic measurements of BPD, HC, AC and FL	SonoBiometry™
US workflow	• Ensures the evaluation of all structures and obtaining the correct planes according to ISUOG guidelines • Comparison between captured images and standardised criteria to ensure quality	ScanNav™ assist SonoLystX™

4. Conclusion

Assessment of the fetus in the second trimester is very important in the management of pregnancy.[6] US is considered the gold standard in pregnancy because of its advantages, but it needs to be improved. Its performance is highly dependent on the sonographer and their manual activity, so automating this examination can mean ameliorate it, and applying AI models is a potential ally in achieving this goal.

Fetal anatomical assessment is one of the most important parts of mid-trimester US, requiring assessment of several structures detailed in ISUOG guidelines. For each structure to be assessed, the sonographer needs to secure multiple planes to ensure that it is normal or to report anatomical changes. Models already developed have demonstrated a high ability to identify and segment different anatomical structures, but the greatest benefit comes from detecting congenital anomalies, allowing early intervention, and reducing fetal and maternal morbidity and mortality. A large amount of data with these anomalies is required to enable automated models to detect those across different organs and systems. However, the number of images reporting them is limited due to their low prevalence in general population. One way to overcome this may be the one presented in the model by *Montero et al.*[32] This model makes it possible to manipulate real images used to generate new artificial images. This increases the database used to teach and train the automatic models.

The heart and CNS are the main anatomical landmarks to which AI models have been applied. This is because the most common congenital abnormalities are found in these anatomical systems. Regarding fetal cardiac evaluation, automatic models allow the detection of various planes indicated for CHDs diagnosis. *Arnaout et al.*, after adding sonographic images of CHDs to the databases they used for their model, were able to demonstrate diagnostic improvement over what is currently seen. In addition, the authors were able to establish the variation of cardiac axis between healthy hearts and CHDs.[3] This example can demonstrate how these models can help on the practical clinical setting.

While targeting the different structures of the CNS is also a major focus of AI, with good accuracies and results, there are other details worth mentioning. For example, *Hesse et al.* used this technology to generate normal growth curves for CNS structures during mid-trimester.[33] The collection of large amounts of data and the use of automated models may allow more rapid determination of reference ranges for structures that have not yet been defined for each gestational age. It could therefore become a new method for determining GA, especially in more advanced pregnancies.

There are certain anatomy regions where the models developed only recognize and segment the levels in which they are found. This is the case for face, thorax, genitalia and abdomen.[35–

[44] According to ISUOG, anomalies in these fetal components should be looked for in mid-trimester, but the automated models described have not been tested. It would therefore be important to apply or improve these models in a real clinical setting to achieve this important topic on the fetal mid-trimester anatomy assessment. In addition, the ISUOG guidelines include other fetal assessment criteria for which there are still no AI models designed for their assessment. It could be important to apply existing or new models to these parameters.

Diagnosis of fetal weight alterations is also important at this stage of pregnancy. Although AI models for estimating fetal weight are still in their early stages, several models are already being applied to improve measurements in fetal biometry.[48,52,53] They report that the measurements obtained automatically are very close to the measurements taken by experienced clinicians, with a low level of associated error. This may indirectly give more accurate results for fetal weight, but more research should be done in this area. In addition, automatically obtaining fetal biometry shows less error associated with GA determination during mid-trimester compared to manually determination by experts.[13,50] We should also mention the model of *Burgos-Artizzu et al.*, which combined fetal biometry with brain texture analysis.[49] The comparison between this method and the use of biometry alone showed that the former was able to obtain a lower error in GA determination. Although it is preferable to calculate it earlier rather than later in pregnancy, these methods of estimating GA can help pregnant women who seek medical care late in pregnancy.

The morphological evaluation of the placenta reflects fetal development and growth and whether the placenta is adequate. The automated models developed were designed to determine placental maturity and were able to achieve high sensitivity, specificity and accuracy.[14,55] Although models already exist to automatically locate the placenta, its relationship to the internal cervical os is sometimes not represented.[58] The development of AI models that can establish the relationship between the position of the placenta and the internal cervical os will be most promising for the second trimester, as they will help to establish this important assessment parameter according to guidelines. Models that aid in the diagnosis of abnormally invasive placenta according to the detection of characteristic lesions have also been mentioned.[57] However, the results obtained are still insufficient and further studies are needed to improve the automatic diagnosis of this pathological situation.

Amniotic fluid assessment, which allows understanding whether its quantity is normal or not, showed very low measurement errors using an automatic model, overlapping those obtained by sonographers.[59,60] One of the automatic models, developed by *Sun et al.*, has the ability to remove noise from the image, something that can potentially make the exam difficult and time-consuming.[60] A similar technique was used to segment the placenta and gave good results by removing artefacts from the image.[54] Potentially, these technologies could have worldwide application in obstetric US, allowing for clearer and more objective images, thus eliminating one of its major disadvantages.

Finally, the analysis of US techniques and workflow was also possible through AI models. This can be essential to assist in creating better models, following the best practice performed by clinicians. On the other hand, these models can also be important to define the best image sequences to analyse, ensuring a complete examination. Its application can mean less workload while performing US scans, allowing less clinician exhaustion and better day-to-day performance. Furthermore, several models report the ability to process and interpret US images within seconds countless times without the need for breaks.[31,38,40,43,46,50,58] This is a clear advantage over clinicians, who are susceptible to fatigue from repetitive tasks. As a result, the time taken to perform the examination can be reduced, allowing for a better doctor-patient relationship and more time to communicate the results and discuss possible therapies. In this respect, AI appears as an adjunct to the doctor´s work.

The prospective study by Mathew *et al.* on AI application to a real clinical setting seems to be beneficial in daily use by sonographers.[63] However, there are still parameters that, despite being close to the manual method, need to be improved. More prospective studies should therefore be the next step in understanding the impact of this technology in the day-to-day work of sonographers and improving the quality of results. Further studies are needed to ensure data privacy of all fetuses and pregnant women in whom this computer technology is used. Although these models are very promising in theory, their use in practice will always need to be fine-tuned and could be the next big step in the improvement of obstetric health care throughout the pregnancy.

In conclusion, the use of AI models seems to be promising, as they allow a better quality of diagnosis of congenital anomalies, a better assessment of structures that ensure fetal growth and health, as well as a more efficient turnaround time of US scans with less clinician fatigue. However, this technology still needs adjustment to real clinical scenarios to fully achieve its potential in helping clinicians worldwide.

5. References

1. Abramowicz JS. Obstetric ultrasound: where are we and where are we going? Ultrasonography. 2021 Jan 1;40(1):57–74.
2. Sharma H, Drukker L, Chatelain P, Droste R, Papageorghiou AT, Noble JA. Knowledge representation and learning of operator clinical workflow from full-length routine fetal ultrasound scan videos. Med Image Anal. 2021 Apr;69:101973.
3. Arnaout R, Curran L, Zhao Y, Levine JC, Chinn E, Moon-Grady AJ. An ensemble of neural networks provides expert-level prenatal detection of complex congenital heart disease. Nat Med. 2021;27(5):882–91.
4. Shu X, Chang F, Zhang X, Shao C, Yang X. ECAU-Net: Efficient channel attention U-Net for fetal ultrasound cerebellum segmentation. Biomed Signal Process Control. 2022 May;75:103528.
5. Zhang B, Liu H, Luo H, Li K. Automatic quality assessment for 2D fetal sonographic standard plane based on multitask learning. Medicine. 2021 Jan 29;100(4):e24427.
6. Salomon LJ, Alfirevic Z, Berghella V, Bilardo CM, Chalouhi GE, da Silva Costa F, et al. <scp>ISUOG</scp> Practice Guidelines (updated): performance of the routine mid-trimester fetal ultrasound scan. Ultrasound in Obstetrics & Gynecology. 2022 Jun 20;59(6):840–56.
7. Afonso AD. DIREÇÃO-GERAL DA SAÚDE Exames Ecográficos na Gravidez de baixo risco Médicos do Sistema Nacional de Saúde [Internet]. 2011. Available from: www.dgs.pt
8. Burgos-Artizzu XP, Coronado-Gutiérrez D, Valenzuela-Alcaraz B, Bonet-Carne E, Eixarch E, Crispi F, et al. Evaluation of deep convolutional neural networks for automatic classification of common maternal fetal ultrasound planes. Sci Rep. 2020;10(1):10200.
9. AIUM Practice Parameter for the Performance of Detailed Second- and Third-Trimester Diagnostic Obstetric Ultrasound Examinations. Journal of Ultrasound in Medicine. 2019 Dec 17;38(12):3093–100.
10. Grandjean H, Larroque D, Levi S. The performance of routine ultrasonographic screening of pregnancies in the Eurofetus Study. Am J Obstet Gynecol. 1999 Aug;181(2):446–54.
11. Nurmaini S, Rachmatullah MN, Sapitri AI, Darmawahyuni A, Tutuko B, Firdaus F, et al. Deep Learning-Based Computer-Aided Fetal Echocardiography: Application to Heart Standard View Segmentation for Congenital Heart Defects Detection. Sensors. 2021 Nov 30;21(23):8007.

12. Bilardo CM, Chaoui R, Hyett JA, Kagan KO, Karim JN, Papageorghiou AT, et al. ISUOG Practice Guidelines (updated): performance of 11–14-week ultrasound scan. Ultrasound in Obstetrics & Gynecology. 2023 Jan 3;61(1):127–43.
13. Namburete AIL, Stebbing R V, Kemp B, Yaqub M, Papageorghiou AT, Alison Noble J. Learning-based prediction of gestational age from ultrasound images of the fetal brain. Med Image Anal. 2015 Apr;21(1):72–86.
14. Lei B, Yao Y, Chen S, Li S, Li W, Ni D, et al. Discriminative Learning for Automatic Staging of Placental Maturity via Multi-layer Fisher Vector. Sci Rep. 2015 Jul 31;5:12818.
15. Drukker L, Noble JA, Papageorghiou AT. Introduction to artificial intelligence in ultrasound imaging in obstetrics and gynecology. Ultrasound Obstet Gynecol. 2020;56(4):498–505.
16. Shinde PP, Shah S. A Review of Machine Learning and Deep Learning Applications. In: 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA). IEEE; 2018. p. 1–6.
17. Currie G, Hawk KE, Rohren E, Vial A, Klein R. Machine Learning and Deep Learning in Medical Imaging: Intelligent Imaging. J Med Imaging Radiat Sci. 2019 Dec;50(4):477–87.
18. Samuel AL. Some Studies in Machine Learning Using the Game of Checkers. IBM J Res Dev. 1959 Jul;3(3):210–29.
19. Yaqub M, Sleep N, Syme S, Chen Z, Ryou H, Walton S, et al. 491 ScanNav® audit: an AI-powered screening assistant for fetal anatomical ultrasound. Am J Obstet Gynecol. 2021 Feb;224(2):S312.
20. Batta Mahesh. Machine Learning Algorithms - A Review. International Journal of Science and Research (IJSR). 2020;9(1):381–6.
21. Erickson BJ, Korfiatis P, Akkus Z, Kline TL. Machine Learning for Medical Imaging. Radiographics. 2017;37(2):505–15.
22. Ray S. A Quick Review of Machine Learning Algorithms. In: 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon). IEEE; 2019. p. 35–9.
23. Nurmaini S, Rachmatullah MN, Sapitri AI, Darmawahyuni A, Tutuko B, Firdaus F, et al. Deep Learning-Based Computer-Aided Fetal Echocardiography: Application to Heart Standard View Segmentation for Congenital Heart Defects Detection. Sensors. 2021 Nov 30;21(23):8007.
24. Yeo L, Romero R. Fetal Intelligent Navigation Echocardiography (FINE): a novel method for rapid, simple, and automatic examination of the fetal heart. Ultrasound Obstet Gynecol. 2013 Sep;42(3):268–84.

25. Carvalho J, Allan L, Chaoui R, Copel J, DeVore G, Hecher K, et al. ISUOG Practice Guidelines (updated): sonographic screening examination of the fetal heart. Ultrasound in Obstetrics & Gynecology. 2013 Mar;41(3):348–59.
26. Dozen A, Komatsu M, Sakai A, Komatsu R, Shozu K, Machino H, et al. Image Segmentation of the Ventricular Septum in Fetal Cardiac Ultrasound Videos Based on Deep Learning Using Time-Series Information. Biomolecules. 2020;10(11).
27. Shozu K, Komatsu M, Sakai A, Komatsu R, Dozen A, Machino H, et al. Model-Agnostic Method for Thoracic Wall Segmentation in Fetal Ultrasound Videos. Biomolecules. 2020 Dec 17;10(12):1691.
28. Huang R, Xie W, Alison Noble J. VP-Nets : Efficient automatic localization of key brain structures in 3D fetal neurosonography. Med Image Anal. 2018;47:127–39.
29. Siddesh A, Gupta G, Sharan R, Agarwal M, Phadke SR. Spectrum of prenatally detected central nervous system malformations: Neural tube defects continue to be the leading foetal malformation. Indian J Med Res. 2017 Apr;145(4):471–8.
30. Malinger G, Paladini D, Haratz KK, Monteagudo A, Pilu GL, Timor-Tritsch IE. Practice Guidelines (updated): sonographic examination of the fetal central nervous system. Part 1: performance of screening examination and indications for targeted neurosonography. Ultrasound in Obstetrics & Gynecology. 2020 Sep;56(3):476–84.
31. Xie HN, Wang N, He M, Zhang LH, Cai HM, Xian JB, et al. Using deep-learning algorithms to classify fetal brain ultrasound images as normal or abnormal. Ultrasound Obstet Gynecol. 2020;56(4):579–87.
32. Montero A, Bonet-Carne E, Burgos-Artizzu XP. Generative Adversarial Networks to Improve Fetal Brain Fine-Grained Plane Classification. Sensors. 2021 Nov 29;21(23):7975.
33. Hesse LS, Aliasi M, Moser F, Haak MC, Xie W, Jenkinson M, et al. Subcortical segmentation of the fetal brain in 3D ultrasound using deep learning. Neuroimage. 2022 Jul;254:119117.
34. Merz E, Pashaj S. Anomalies of the Fetal Face. Donald School Journal of Ultrasound in Obstetrics and Gynecology. 2019 Mar 1;13(1):34–40.
35. Wang X, Liu Z, Du Y, Diao Y, Liu P, Lv G, et al. Recognition of Fetal Facial Ultrasound Standard Plane Based on Texture Feature Fusion. Comput Math Methods Med. 2021;2021:6656942.
36. Lei B, Tan EL, Chen S, Zhuo L, Li S, Ni D, et al. Automatic Recognition of Fetal Facial Standard Plane in Ultrasound Image via Fisher Vector. PLoS One. 2015 May 1;10(5):e0121838.

37. Yu Z, Ni D, Chen S, Li S, Wang T, Lei B. Fetal facial standard plane recognition via very deep convolutional networks. In: 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE; 2016. p. 627–30.
38. Tsai PY, Chen HC, Huang HH, Chang CH, Fan PS, Huang CI, et al. A new automatic algorithm to extract craniofacial measurements from fetal three-dimensional volumes. Ultrasound in Obstetrics & Gynecology. 2012 Jun;39(6):642–7.
39. Rahmatullah B, Sarris I, Noble JA, Papageorghiou AT. OP26.08: Automated standard plane selection from fetal abdominal ultrasound volumes using a machine learning algorithm. Ultrasound in Obstetrics & Gynecology. 2012 Sep;40(S1):134–5.
40. Ahmed M, Noble JA. An eye-tracking inspired method for standardised plane extraction from fetal abdominal ultrasound volumes. In: 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI). IEEE; 2016. p. 1084–7.
41. Cai Y, Sharma H, Chatelain P, Noble JA. SonoEyeNet: Standardized fetal ultrasound plane detection informed by eye tracking. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). IEEE; 2018. p. 1475–8.
42. Maysanjaya IMdD, Nugroho HA, Setiawan NA. The classification of fetus gender on ultrasound images using learning vector quantization (LVQ). In: 2014 Makassar International Conference on Electrical Engineering and Informatics (MICEEI). IEEE; 2014. p. 150–5.
43. Tang S, Chen S ping. A fast automatic recognition and location algorithm for fetal genital organs in ultrasound images. J Zhejiang Univ Sci B. 2009 Sep 30;10(9):648–58.
44. Kaplan E, Ekinci T, Kaplan S, Barua PD, Dogan S, Tuncer T, et al. PFP-LHCINCA: Pyramidal Fixed-Size Patch-Based Feature Extraction and Chi-Square Iterative Neighborhood Component Analysis for Automated Fetal Sex Classification on Ultrasound Images. Contrast Media Mol Imaging. 2022 May 18;2022:1–10.
45. Lees CC, Stampalija T, Baschat AA, Silva Costa F, Ferrazzi E, Figueras F, et al. ISUOG Practice Guidelines: diagnosis and management of small-for-gestational-age fetus and fetal growth restriction. Ultrasound in Obstetrics & Gynecology. 2020 Aug;56(2):298–312.
46. Ambroise Grandjean G, Hossu G, Bertholdt C, Noble P, Morel O, Grangé G. Artificial intelligence assistance for fetal head biometry: Assessment of automated measurement software. Diagn Interv Imaging. 2018 Nov;99(11):709–16.
47. Papageorghiou AT, Kemp B, Stones W, Ohuma EO, Kennedy SH, Purwar M, et al. Ultrasound-based gestational-age estimation in late pregnancy. Ultrasound Obstet Gynecol. 2016 Dec;48(6):719–26.

48. Cai Y, Droste R, Sharma H, Chatelain P, Drukker L, Papageorghiou AT, et al. Spatio-temporal visual attention modelling of standard biometry plane-finding navigation. Med Image Anal. 2020 Oct;65:101762.
49. Burgos-Artizzu XP, Coronado-Gutiérrez D, Valenzuela-Alcaraz B, Vellvé K, Eixarch E, Crispi F, et al. Analysis of maturation features in fetal brain ultrasound via artificial intelligence for the estimation of gestational age. Am J Obstet Gynecol MFM. 2021;3(6):100462.
50. Dan T, Chen X, He M, Guo H, He X, Chen J, et al. DeepGA for automatically estimating fetal gestational age through ultrasound imaging. Artif Intell Med. 2023 Jan;135:102453.
51. Yang C, Yang Z, Liao S, Guo J, Yin S, Liu C, et al. A new approach to automatic measure fetal head circumference in ultrasound images using convolutional neural networks. Comput Biol Med. 2022 Aug;147:105801.
52. Wang X, Wang W, Cai X. Automatic measurement of fetal head circumference using a novel GCN-assisted deep convolutional network. Comput Biol Med. 2022 Jun;145:105515.
53. Zhu F, Liu M, Wang F, Qiu D, Li R, Dai C. Automatic measurement of fetal femur length in ultrasound images: a comparison of random forest regression model and SegNet. Mathematical Biosciences and Engineering. 2021;18(6):7790–805.
54. Hu R, Singla R, Yan R, Mayer C, Rohling RN. Automated Placenta Segmentation with a Convolutional Neural Network Weighted by Acoustic Shadow Detection. In: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE; 2019. p. 6718–23.
55. Li X, Yao Y, Ni D, Chen S, Li S, Lei B, et al. Automatic staging of placental maturity based on dense descriptor. Biomed Mater Eng. 2014;24(6):2821–9.
56. Qi H, Collins S, Noble A. Weakly Supervised Learning of Placental Ultrasound Images with Residual Networks. In 2017. p. 98–108.
57. Qi H, Collins S, Noble JA. Automatic Lacunae Localization in Placental Ultrasound Images via Layer Aggregation. Med Image Comput Comput Assist Interv. 2018 Sep;11071:921–9.
58. Schilpzand M, Neff C, van Dillen J, van Ginneken B, Heskes T, de Korte C, et al. Automatic Placenta Localization From Ultrasound Imaging in a Resource-Limited Setting Using a Predefined Ultrasound Acquisition Protocol and Deep Learning. Ultrasound Med Biol. 2022 Apr;48(4):663–74.
59. Cho HC, Sun S, Min Hyun C, Kwon JY, Kim B, Park Y, et al. Automated ultrasound assessment of amniotic fluid index using deep learning. Med Image Anal. 2021;69:101951.

60. Sun S, Kwon JY, Park Y, Cho HC, Hyun CM, Seo JK. Complementary Network for Accurate Amniotic Fluid Segmentation From Ultrasound Images. IEEE Access. 2021;9:108223–35.
61. Drukker L, Sharma H, Karim JN, Droste R, Noble JA, Papageorghiou AT. Clinical workflow of sonographers performing fetal anomaly ultrasound scans: deep-learning-based analysis. Ultrasound in Obstetrics & Gynecology. 2022 Dec;60(6):759–65.
62. Alsharid M, Cai Y, Sharma H, Drukker L, Papageorghiou AT, Noble JA. Gaze-assisted automatic captioning of fetal ultrasound videos using three-way multi-modal deep neural networks. Med Image Anal. 2022 Nov;82:102630.
63. Matthew J, Skelton E, Day TG, Zimmer VA, Gomez A, Wheeler G, et al. Exploring a new paradigm for the fetal anomaly ultrasound scan: Artificial intelligence in real time. Prenat Diagn. 2022 Jan 18;42(1):49–59.
64. GE Healthcare. (2021). Voluson Swift Ultrasound Machine. GE Healthcare website. Retrieved September 22, 2021, from https://www.gehealthcare.co.uk/products/ultrasound/voluson/voluson-swift.