NOVA SCHOOL OF SCIENCE & TECHNOLOGY

DEPARTMENT OF PHYSICS

ADRIANA ABREU RAMOS

Bachelor of Science in Biomedical Engineering

SEMI-AUTOMATIC EVALUATION OF FETAL GROWTH RESTRICTION USING ULTRASOUND IMAGES

MASTER IN BIOMEDICAL ENGINEERING

NOVA University Lisbon
March, 2023

NOVA SCHOOL OF SCIENCE & TECHNOLOGY

DEPARTMENT OF PHYSICS

SEMI-AUTOMATIC EVALUATION OF FETAL GROWTH RESTRICTION USING ULTRASOUND IMAGES

ADRIANA ABREU RAMOS

Bachelor of Science in Biomedical Engineering

Adviser: Professor Doutor José Manuel Matos Ribeiro da Fonseca
Associated Professor with Aggregation, NOVA University of Lisbon

Co-adviser: Doutor Alcides Horácio Pereira
Gynecologist and Obstetrician, Hospital Garcia de Orta

MASTER IN BIOMEDICAL ENGINEERING

NOVA University Lisbon
March, 2023

Semi-automatic Evaluation of Fetal Growth Restriction using Ultrasound Images

To my family.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my adviser, Professor José Manuel Fonseca, for allowing me to work with him in an area that I have been interested in for a long time. Thank you for all your support, availability, and encouragement throughout the entire process and for sharing your knowledge and ideas. I will always be grateful for that.

Also, I would like to thank my co-adviser, Doctor Alcides Pereira, and Doctor Catarina Silva for giving me the opportunity to realise my dissertation in the medical field and for all their clinical insights. To Doctor Mariana Silva, thank you for being so tireless. Without your help, none of this would have been possible.

To all the friends this amazing university has given to me, I would also like to express my gratitude for making these last five years the most special years of my entire life. All of you have my heart.

A special thank you to Isa for always being there for me and for being the best friend I could have ever asked for. I will never forget all the times you had to deal with my dramas, the nights we spent together playing board games and so much more.

To my boyfriend, João Freire, thank you for the unconditional love and for believing in me even when I did not. Thank you for always being there to listen and for being the calm that I need in my life.

Lastly, the most sincere thank you to all my family. To my mother and my father, thank you for supporting me unconditionally and for always believing in me no matter what. To my little sister, my eternal soulmate, thank you for so much... Thank you for your patience, for making me laugh and for always picking me up in the most difficult times. And to Inês, who has been and always will be my role model as a woman, thank you for helping me to grow and for making me see life in such a beautiful way.

Abstract

Fetal growth restriction (FGR) is a condition where the developing fetus is unable to achieve the expected growth potential for its gestational age. It is the second leading cause of perinatal mortality and fetuses are more susceptible to the onset of numerous long-term diseases. Thus, early diagnosis and effective monitoring of FGR are very important.

Ultrasound is the most widely used medical imaging technique for the diagnosis and surveillance of growth-restricted fetuses. In current clinical practice, to obtain an estimate of fetal age it is necessary to perform the measurement of four biometric indices. Apart from manually tracing being time-consuming and highly dependent on clinicians, the intra- and inter-observer variability influences the accuracy and consistency of measurements.

Since clinicians believe that the liver of fetuses in the lowest percentiles is one of the most determinant indicators to detect changes in fetal development, this dissertation aimed to propose a novel approach that attempts to correlate the fetal liver echogenicity and the diagnosis of FGR, using image processing techniques and Machine Learning models. One of the advantages of this method is that it can improve the effectiveness of FGR diagnosis by eliminating the need for an exhaustive collection of biometric indices.

To this purpose, a computer-assisted tool was developed which allows manual segmentation of a portion of the fetal liver. Next, a textural analysis of the various segments was performed. Finally, in order to identify the best classification model that distinguishes fetuses with FGR from those with normal fetal growth based on their textural features, five supervised learning models were trained and evaluated. After a comparative analysis, the results showed that the model that produced the best results was the Decision Tree, with an accuracy of 0.77 and an area under the receiver operating characteristic (AUROC) curve of 0.62.

Although the results obtained cannot be considered highly conclusive due to the limited and unrepresentative nature of the available dataset, it is possible to state that the proposed approach is feasible and has the potential to establish a correlation between the echogenicity of the fetal liver and changes in fetal development. However, to arrive at definitive conclusions, it will be necessary to delve deeper into the study that has been initiated, using a larger volume of data.

Keywords: Fetal Growth Restriction, Liver Echogenecity, Ultrasound, Computer-assisted Tool, Artificial Intelligence

Resumo

A restrição do crescimento fetal (RCF) ocorre quando o feto não atinge o potencial de crescimento determinado para a sua idade gestacional. É a segunda principal causa de morte perinatal e os fetos são mais suscetíveis a contrair inúmeras doenças a longo prazo. Desta forma, o diagnóstico precoce e a monitorização eficaz da RCF são muito importantes.

A ecografia é a técnica de imagem médica mais utilizada para o diagnóstico e a vigilância de fetos com um crescimento restrito. Na prática clínica atual, para obter uma estimativa da idade do feto é necessário realizar a medição de quatro índices biométricos. Para além do rastreio manual ser demorado e muito dependente dos clínicos, a variabilidade intra- e inter-observadores influencia a precisão e consistência das medições.

Dado que os clínicos acreditam que o fígado dos fetos nos percentis mais baixos é um dos indicadores mais determinantes para detetar alterações no desenvolvimento fetal, esta dissertação teve como principal objetivo propor uma abordagem inovadora que tenta correlacionar a ecogenicidade do fígado fetal e o diagnóstico da RCF, utilizando técnicas de processamento de imagem e modelos de Aprendizagem Automática. Uma das vantagens deste método é que pode melhorar a eficácia do diagnóstico da RCF ao eliminar a necessidade de uma recolha exaustiva dos índices biométricos.

Nesse sentido, foi desenvolvida uma ferramenta que permite realizar a segmentação manual de uma porção do fígado fetal. Em seguida, foi realizada uma análise textural dos vários segmentos. Por último, de modo a identificar o modelo de classificação que melhor distingue os fetos com RCF dos fetos com um crescimento fetal normal com base nas suas características texturais, foram treinados e avaliados cinco modelos de aprendizagem supervisionada. Após uma análise comparativa, foi possível concluir que o modelo que produziu os melhores resultados foi a Árvore de Decisão, com uma precisão de 0.77 e uma área sob a curva *Receiver Operating Characteristic* (ROC) de 0.62.

Embora os resultados obtidos não se possam considerar muito conclusivos, uma vez que o conjunto de dados disponível era bastante limitado e pouco representativo, é possível afirmar que a abordagem proposta é viável e que apresenta potencial para estabelecer uma relação entre a ecogenicidade do fígado fetal e as alterações no desenvolvimento do feto. No entanto, para obter conclusões definitivas, será necessário aprofundar o estudo agora iniciado, recorrendo a um maior volume de dados.

Palavras-chave: Restrição do Crescimento Fetal, Ecogenecidade Hepática, Ecografia, Ferramenta Assistida por Computador, Inteligência Artificial

Contents

List of Figures

List of Tables

List of Acronyms and Abbreviations

2D Two-dimensional *(pp. 26–28)*

AC Abdominal Circumference *(pp. ix, 6, 8–10, 17, 18, 23–30)*
AI Artificial Intelligence *(pp. 2, 18, 23, 30, 59)*
ANN Artificial Neural Network *(pp. 27, 28)*
ASM Angular Second Moment *(pp. 41, 42, 48, 49, 70)*
AUROC Area Under the Receiver Operating Characteristic *(pp. 22, 55, 56, 58, 60)*

BPD Biparietal Diameter *(pp. 8, 9, 17, 18, 23, 24)*

CNN Convolutional Neural Network *(pp. 26, 28)*
CRL Crown-Rump Length *(pp. 9, 17, 25)*

DICOM Digital Imaging and Communication in Medicine *(p. 30)*
DL Deep Learning *(p. 2)*
DT Decision Tree *(pp. x, 19, 31, 43, 52, 54–56, 58, 60)*

EDA Exploratory Data Analysis *(pp. x, 50, 51, 59)*
EFW Estimated Fetal Weight *(pp. 6, 9, 10)*

FASP Fetal Abdominal Standard Plane *(pp. 26, 28)*
FGR Fetal Growth Restriction *(pp. 1–4, 6–10, 16, 17, 23, 26–30, 42, 48, 50, 51, 53, 55, 56, 59, 60)*
FL Femur Length *(pp. 8–10, 17, 18, 23, 24)*
FN False Negative *(pp. 20, 21, 53, 55, 56)*
FP False Positive *(pp. 20–22, 53, 55, 56)*
FPR False Positive Rate *(p. 22)*

GA Gestational Age *(pp. 1, 7–10, 16, 17, 23)*
GLCM Grey Level Co-occurrence Matrix *(pp. ix, xi, 40, 41, 48, 70)*
GVF Gradient Vector Field *(pp. 25, 26, 28, 30)*

HC Head Circumference *(pp. 8–10, 17, 18, 23–26)*

HGO Hospital Garcia de Orta *(pp. 2, 3, 29, 30, 33, 45, 50, 52, 59)*
HTML HyperText Markup Language *(p. 50)*

IRHT Iterative Randomized Hough Transform *(pp. 25, 26, 28, 30)*

kNN k-Nearest Neighbour *(pp. x, 19, 31, 43, 52, 54–57)*

ML Machine Learning *(pp. xi, 2, 3, 18–20, 22, 27, 28, 31, 42–44, 50, 52, 53, 55, 56, 58, 60)*

NB Naive Bayes *(pp. x, 19, 31, 43, 52, 54–57)*
NN Neural Network *(pp. 19, 26, 28)*

RF Random Forest *(pp. x, 19, 26, 28, 31, 43, 52, 54–56)*
ROC Receiver Operating Characteristic *(pp. ix, x, 22, 56, 57)*
ROI Region of Interest *(pp. ix, x, 24, 26, 30, 34, 37, 45, 46, 59)*

SACM Scale-aware Auto-context Model *(pp. 26, 28)*
SE Structuring Element *(pp. 34, 35)*
SFH Symphysis Fundal Height *(pp. 9, 23)*
SGA Small for Gestational Age *(pp. 6, 7)*
SL Supervised Learning *(pp. 19, 27)*
SVM Support Vector Machine *(pp. x, 19, 31, 43, 52, 54–57)*

TN True Negative *(pp. 20, 21, 53, 55, 56)*
TP True Positive *(pp. 20–22, 53, 55, 56)*
TPR True Positive Rate *(pp. 21, 22)*

UL Unsupervised Learning *(pp. 19, 27)*
US Ultrasound *(pp. ix, 1–3, 9, 11–17, 23–31, 33, 34, 37, 39, 46, 52, 59)*

CHAPTER 1

INTRODUCTION

The research work described in this dissertation was carried out in accordance with the norms established in the ethics code of Universidade Nova de Lisboa. The work described and the material presented in this dissertation, with the exceptions clearly indicated, constitute original work carried out by the author.

1.1 Context and Motivation

Fetal Growth Restriction (FGR), also known as intrauterine growth restriction, is a common pregnancy complication that has been associated with some negative perinatal outcomes. In this pregnancy condition, the fetus does not reach its growth potential determined for its Gestational Age (GA) due to a pathological factor, most commonly placental dysfunction. In practice, FGR is reflected in a decrease in fetal size percentiles that are used to determine fetal weight during pregnancy.

This clinical condition affects about 5 – 10% of pregnancies and varies according to population, geographical location and growth charts used as reference. It is the second leading cause of perinatal mortality and accounts for 30% of stillborn infants. Moreover, FGR is one of the most common causes of premature births and epemiological studies have revealed that fetuses are more susceptible to the onset of neurological, cardiovascular, and endocrine disorders in the long term. Thus, early diagnosis and monitoring of FGR are extremely important.

Ultrasound (US) is the most commonly used medical imaging technique during pregnancy, as it provides valuable information about fetal growth and development, by assessing fetal biometry and the functionality of certain organs. Moreover, it is a noninvasive technique which is very useful in the detection of fetal growth abnormalities, such as FGR.

However, manual screening of US images is not only time-consuming but also highly dependent on the expertise of the clinician. Intra- and interobserver variability can further impact the accuracy and consistency of measurements, leading to potential errors and misdiagnosis. Therefore, there is a growing need for more reliable and automated methods for fetal biometry analysis that can reduce the burden on clinicians and improve the accuracy and consistency of measurements.

Over the last few years, researchers have developed Artificial Intelligence (AI)-based methods to overcome the aforementioned limitations. By automatically analysing US images, these methods have the potential to reduce misdiagnosis and missed diagnosis rates, thereby improving medical outcomes. Besides the application of AI in the fields of obstetrics and gynaecology is still in the early stages, studies have shown that the implementation of Machine Learning (ML) and Deep Learning (DL) models has been showing promising results in pattern recognition and automatic extraction of biometric indices from US images [1]. While many of these models and other AI solutions can provide significant benefits in healthcare, it's crucial to note that some of them require human expertise and oversight. This is because fully automated solutions may not always be appropriate or effective in clinical settings, and unexpected events may require specialist intervention. Therefore, it's important for healthcare professionals to work closely with AI tools and ensure that they are used safely and effectively to improve patient outcomes.

1.2 Thesis Goals

This thesis arises from the need expressed by the Hospital Garcia de Orta (HGO) [1] to enhance the diagnosis of FGR through the development of an automated method using US images. The focus of this method is to analyse the echogenicity of the fetal liver, as clinicians consider it a crucial indicator for detecting changes in fetal development. While existing methods have focused on distinguishing various types of FGR, there is currently no known approach that explores the relationship between liver echogenicity and the diagnosis of this medical condition.

Therefore, the aim of this dissertation is to develop a computer-assisted tool that employs AI models and basic image processing techniques to segment and analyse the fetal liver in order to establish a potential correlation between hepatic echogenicity and the diagnosis of FGR. The proposed method encompasses several steps, including managing and processing the image dataset, developing and refining ML algorithms, and ultimately evaluating the model's performance in predicting fetuses with growth restriction.

[1] HGO, located in Almada, is a prominent public hospital recognised as one of the leading healthcare institutions in Portugal. Renowned for its specialised departments and diverse medical expertise, this expansive facility delivers a wide range of medical services. HGO is a reference hospital for the surrounding region, offering both primary and specialised healthcare services to the local population.

Regarding clinical contributions, the development of this novel approach offers gynaecologists and obstetricians a tool that enables faster and more accurate diagnosis. This advancement leads to reduced examination time, improved workflow, and optimised pregnancy follow-up plans of growth-restricted fetuses.

1.3 Thesis Outline

This document is structured into six Chapters. The first Chapter provides an introduction to the dissertation, including its motivation, main objectives, and an overview of the document's contents. The second Chapter contains all the theoretical concepts required for the development of this dissertation, including normal fetal growth and development, a contextualisation of FGR and US imaging, and an overview of ML. The third chapter presents a literature review of FGR diagnosis, focusing on state-of-the-art manual, automatic and semi-automatic methods.

The fourth Chapter covers the information about the dataset from HGO and a comprehensive description of all the methods and techniques adopted to develop the proposed work. The fifth Chapter presents a critical analysis of the results obtained. Finally, the sixth Chapter concludes the present dissertation where a brief summary of the study and the main findings are present, as well as limitations and suggestions for future research directions.

CHAPTER 2

THEORETICAL CONCEPTS

2.1 Fetal Growth and Development

Fetal growth and development concerns to fetus growth in terms of size and it can be assessed through anthropometric measurements and changes in body composition such as the amount of fat and muscle [2]. It is influenced by physiological, metabolic and hormonal adaptions (e.g. minimisation of metabolic reliance on glucose and improvement of amino acid oxygenation).

Numerous studies have found that placental insufficiency contributes to the etiology of FGR because it is one of the underlying reasons of changes in fetal development and impaired placental ability to transfer nutrients (e.g. aminoacids and glucose) to the fetus. Indeed, lower expression and/or activity of placental nutrition transporters have been linked to a number of disorders known to increase the chance of having a small or growth-restricted infant [3].

Thus, fetal growth assessment is crucial for the clinicians because changes in fetal growth are mostly associated with the high rates of mortality and morbidity during the prenatal period. Furthermore, it may also act as an important precedent with regard to both childhood and adult diseases.

2.1.1 Normal Fetal Growth

Normal fetal growth and development of the fetus can be divided in three different phases. The first phase, also known as cellular hyperplasia phase, occurs in the first 16 weeks of pregnancy and comprises cell replication and proliferation. Between 16 and 32 weeks, there is a second phase of cell migration and aggregation to form different tissues and primordial organs. Finally, from 32 week of pregnancy, there is a third phase of a rapid increase in cell size and formation of functional organs, also known as cellular hypertrophy phase [4].

Therefore, in the first phase, where there is a very high mitotic activity, fetal growth is mainly influenced by genetic factors, whereas hormonal and environmental factors are more prevalent in the second and third phases.

2.1.2 Placental Nutrient Transport

Placenta is an essential organ for the maintenance of a healthy, viable, and optimally growing fetus. As well as mediating exchange of nutrients, oxygen and waste products between mother and fetus, it acts as a physical barrier that affords some protection to fetus against pathogens and has the ability to synthesise and secrete hormones, growth factors and cytokines that have an important role in the fetal growth [4].

The transfer of substances between maternal and fetal circulations is influenced by uteroplacental and umbilical blood flow, concentration gradients of nutrients and thickness, exchange area and metabolism of the placenta. While the transfer of membrane permeable molecules (e.g., oxygen and carbon dioxide) is influenced by blood flow and placental structure, the transfer of larger and less membrane permeable substances requires transporter-proteins to cross the placenta, in a process that depends on ionic gradients.

Glucose

Carbohydrate is the fetus' major fuel, accounting for around 80% of fetal energy intake [5]. In particular, glucose is the primary energy substrate for the fetus.

Placental glucose transport occurs in a constant uninterrupted way during pregnancy, by facilitated diffusion through specific glucose transporter proteins expressed in both plasma membranes. As glucose levels are higher in the maternal circulation than in the fetal circulation, there is a gradient that favour the transfer of glucose to the fetus.

Placental supply of glucose to the fetus increases throughout pregnancy due to increased fetal glucose transporters and decreased of fetal glucose levels due to increase in insulin dependent uptake of glucose by adipose tissue and skeletal muscle.

Alternate Fuels

Figure 2.1 shows that fetus is capable of using other substrates as fuel. Approximately 40−50% of the supplied glucose is oxidised or converted to glycogen and lactate by the placenta. This lactate, coupled with glucose, is used by the fetus to synthesise fat and glycogen [5].

Glycogen

Some of the energy and substrate available in the third trimester of pregnancy are used to meet fetal growth and development needs as well as to store energy.

Although glycogen is produced by the brain, heart, lungs and skeletal muscle during this time of pregnancy, liver glycogen storage is the most important carbohydrate reserve for systemic glucose demands soon after birth [2]. Additionally, fetal liver is the organ that contains the whole set of enzymes essential for the synthesis storage, followed by heart, and the only one that has enough glucose-6-phosphatase to release glucose in the circulation.

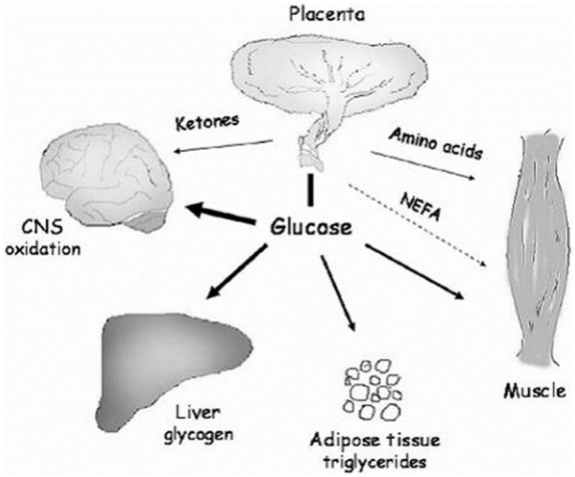

Figure 2.1: Transported and deposited energy substrates in the fetus [5].

Glycogen synthase and glycogen phosphorylase are two enzymes whose functional balance is responsible for the production or degradation of glycogen. So, when glycogen synthase is activated via insulin, prenatal hyperglycemia enhances fetal glycogen deposition, whereas when glycogen phosphorylase is activated, prenatal hypoglycemia stimulate glucose release from the liver.

At the end of normal pregnancies, the fetal liver glycogen content is at least twice that of an adult. In FGR cases, hepatic glycogen stores are lower. This is related to placental insufficiency and low glucose supply rates because, in these cases, metabolic fuels will be used for growth and not for glycogen synthesis [2], [5].

2.2 Fetal Growth Restriction

FGR is a common pregnancy complication, which occurs when there is a suboptimal fetal environment that prevents the fetus from reaching its intrauterine growth potential due to a variety of factors. A compromise in placental function is the most common factor.

Since there is still no gold standard for the diagnosis of FGR, the task of identifying growth restricted fetuses in clinical settings is difficult. It is usually defined by a drop in fetal size percentiles during pregnancy which requires serial assessments that are usually not available [6]. Therefore, clinicians most commonly use the term Small for Gestational Age (SGA) to suspect FGR.

SGA fetuses encompasses fetuses that are constitutionally small but healthy at lower risk of abnormal perinatal outcome. In clinical practice, they are fetuses with an Estimated Fetal Weight (EFW) or Abdominal Circumference (AC) which lies below a certain threshold such as the 10^{th} percentile [6]. Nevertheless, it is important to acknowledge that there are various limitations to utilising SGA as a substitute for FGR.

As mentioned above, SGA fetuses encompasses small healthy fetuses, whose smallness is simply due to their predetermined growth potential (i.e. false-positive diagnosis of FGR). In

fact, approximately 70% of fetuses have a weight below the 10th percentile and this is only due to constitutional factors (e.g. female sex, ethnicity of parents) [4]. On the other hand, some growth-restricted fetuses who remains above the percentile previously described may not reach their original growth potential. Consequently, they may remain non-diagnosed despite being at increased risk of adverse outcomes (i.e. false-negative diagnosis of FGR) [6]. Finally, the diagnosis of SGA is affected by the poor accuracy of sonographic fetal weight estimation and the growth chart being used. All of this highlights the need for a clear definition of both FGR and SGA[1].

In an attempt to overcome some of the limitations described above and decrease the likelihood of false-positive and false-negative diagnosis of FGR, in 2016 a panel of experts joined to reach a consensus-based definition for placenta-mediated FGR through a Delphi procedure [7]. The consensus definition was based on a combination of biometric and functional parameters where the 32nd week of gestation was considered the most suitable threshold to distinguish both early-onset FGR(< 32 weeks) and late-onset FGR (≥ 32 weeks), as described in Figure 2.2.

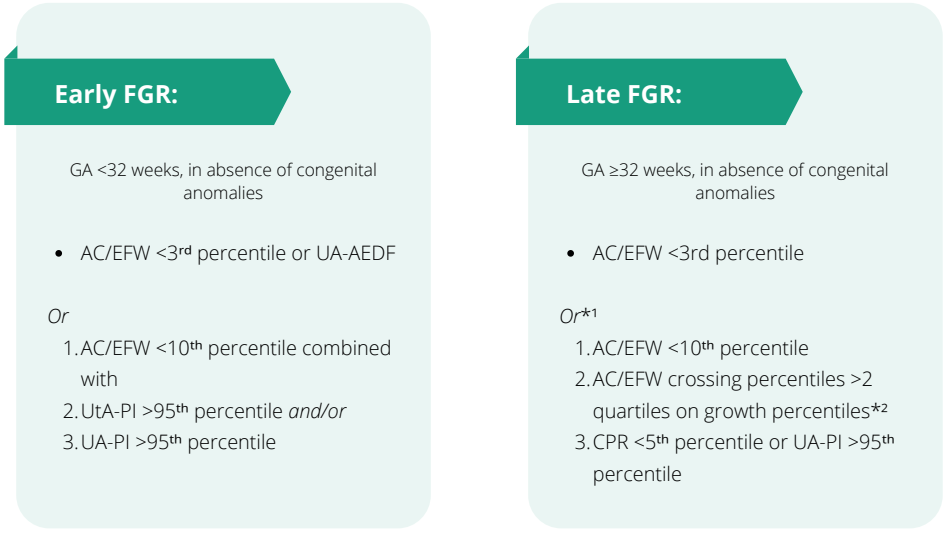

Figure 2.2: Definitions of early and late growth restriction established through the Delphi procedure.

[1]Although there is some inconsistency in the literature regarding the definitions of FGR and SGA, for the purposes of this master's thesis, the term **FGR** will refer to a small fetus that has failed to achieve its growth potential due to a pathological process, while the term **SGA** will refer to an infant with a birth weight below the 10th percentile for their GA.

2.2.1 Classification

According to Campbell [8], fetuses with FGR should be differentiated according to the ratio between the Head Circumference (HC) and the AC (HC/AC), and classified into three types: asymmetrical FGR, symmetrical FGR and mixed FGR.

However, the chronological categorisation of FGR is the most generally used classification at the moment. According to what was previously mentioned, it has been suggested that FGR should be categorised into early-onset FGR ($<$ 32 weeks) and late-onset FGR ($\geq$ 32 weeks), depending on fetus' GA at the time of diagnosis [7].

2.2.1.1 Based on fetal biometry

Asymmetrical FGR

Approximately 70 − 80% of FGR cases exhibit characteristics of asymmetric growth restriction. This type of FGR occurs later, in third trimester of gestation, and it is commonly associated with placental insufficiency and redistribution of fetal blood to vital organs. Furthermore, fetuses have a reduced cell size and normal cell number [9], [10].

Although fetuses show normal growth until the third trimester of gestation, they have a poor weight gain after that. So, fetal weight is the most affected parameter. Antenatal scan demonstrate decreased AC which is mostly due to decreased glycogen storage, smaller liver size, and decreased subcutaneous fat [10]. Additionally, it reveals normal Biparietal Diameter (BPD), HC and Femur Length (FL) [9], [10].

Symmetrical FGR

This type of FGR accounts for about 20 − 30% of total FGR cases. It is an early-onset condition where fetuses have a reduced cell number due to genetic disorders or congenital infections. Antenatal scan demonstrate that all biometric measures are affected. There is a proportional decrease in BPD, HC, AC and FL [9], [10]. When compared with asymmetrical FGR fetuses in terms of mortality and morbidity, fetuses with symmetrical FGR have a poor prognosis because they are at greater risk of adverse outcomes [9], [11].

2.2.1.2 Based on diagnosis time

Early-onset FGR

Early-onset FGR has a prevalence of 0.5 − 1% and is typically more severe [6]. It tends to present an established Doppler changes pattern in the umbilical artery and ductus venosus and it is more commonly associated with cases of maternal hypertensive disorders and placental dysfunction [7], [12]–[15]. Early-onset FGR diagnosis tends to be more easier than diagnosis of late-onset FGR and the main difficulty of these cases is management (i.e. timing of delivery).

Late-onset FGR

Late-onset FGR has a prevalence of 5 − 10% so it is more common than early-onset FGR [6]. In these cases, it is more usual to observe a normal umbilical artery Doppler and diagnosis

is the main challenge. It is based on changes in cerebral circulation that result in a low flow resistance in the middle cerebral artery, leading to a low cerebroplacental ratio [16], [17].

Regarding to management, this is more easier than in early-onset FGR cases because it is usually made in final stages of pregnancy, where the risks associated with delivery are minor [6].

2.2.2 Etiology

FGR is the common end result of countless fetal, maternal or placental disorders that interfere with the mechanisms that regulate normal fetal growth. It can also result due to a combination of any of the factors mentioned above.

Various fetal disorders such as chromosomal abnormalities (e.g. tisomy 13, 18 and 21), genetic syndromes, congenital anomalies (e.g. congenital heart disease), multiple gestation, congenital infections (e.g. toxoplasmosis, cytomegalovirus and malaria) and metabolic disorders are responsible for causing FGR [4], [18]. On the other hand, there is a lot of maternal factors that are responsible for causing it too. Maternal medical conditions that may result in FGR include hypertensive disorders, diabetes associated with vasculopathy or chronic renal disease. Additionally, inter-pregnancy interval, maternal health, behavioural habits (e.g. maternal substance abuse which may include smoking, alcohol and illicit drugs), maternal infections and demographic factors like mother's age, low socioeconomic status, ethnicity or race can also affect the healthy growth of the fetus and lead to FGR [18].

As mentioned in 2.1.2, the placenta is an organ with an important role during pregnancy. Any mismatch between the placenta's supply and the fetus's demand will affect the normal growth of the fetus. Abnormal uteroplacental vasculature, placental dysfunction, velamentous cord insertion, single umbilical artery and confined placental mosaicism are some of placental causes of FGR [18].

2.2.3 Diagnosis

Early and proper diagnosis of FGR is crucial because it allows for adequate follow-up of fetuses who are at risk of adverse perinatal outcomes. Since FGR diagnosis is based on the identification of fetuses who are smaller than expected for their GA, maternal and family history, physical examination (i.e. maternal pre-pregnancy weight and height, nutritional status, Symphysis Fundal Height (SFH) and fetal palpation) or US evaluation is needed [6].

The most important parameters for FGR diagnosis is the correct determination of GA and EFW, since these are relevant for tracking fetal growth [18]. GA can be assessed using the date of the last menstrual period but it is often inaccurate [19]. Therefore, measuring the Crown-Rump Length (CRL) during a first-trimester US is the most precise way to date a pregnancy and establishes GA with an accuracy of 5 days in 95% of cases [20]–[22]. EFW can be assessed trough the use of regression equations that combine four biometric indices measured in second-trimester US: AC, BPD, FL and HC [23]–[25]. Many equations have been published, but a recent systematic review [25] discovered that equations based on $3-4$ biometric indices provide the

most accurate results. Hadlock equation based on AC, FL and HC provides the best estimation result of fetal weight (Equation 2.1).

$$\log_{10} EFW = (1.326 - 0.00326 \times AC \times FL + 0.0107 \times HC + 0.0438 \times AC + 0.158 \times FL)^2 \quad (2.1)$$

However, if EFW is below the 10^{th} percentile for GA, complementary diagnostic tools such as Doppler blood flow studies (specially in umbilical artery) and amniotic fluid assessment should be adopted [26]. Biophysical profile, which reflects fetal acid-base status, is another technique that can be used to monitor fetuses that have already been diagnosed with FGR.

2.2.4 Consequences of Fetal Growth Restriction

The main short and long-term consequences of FGR are associated with fetal and obstetric complications. Stillbirth is the most devastating complication [27], [28] and FGR infants have also a higher risk of preterm delivery.

Despite improvements in neonatal care, neonatal mortality and short-term morbidity rates remain higher. Then, FGR infants are prone to present some immediate neonatal complications such as asphyxia, hypothermia, hypoglycaemia, hyperglycaemia, hypocalcaemia, polycythaemia, pulmonary haemorrhage, jaundice, necrotizing enterocolitis, late onset sepsis and so on (see Figure 2.3) [29].

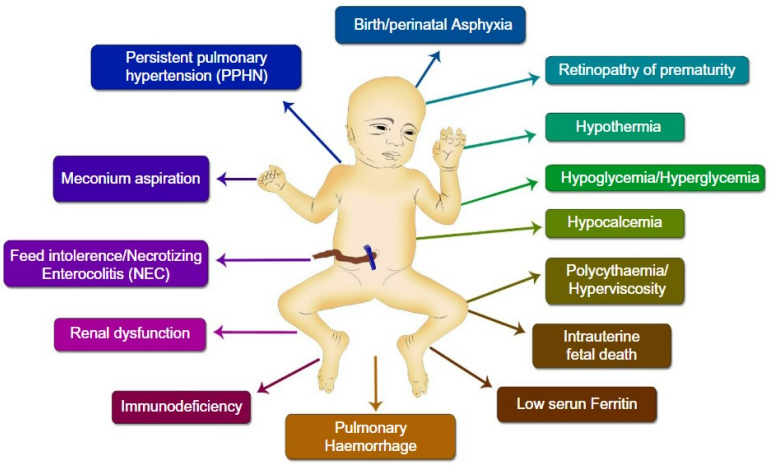

Figure 2.3: Neonatal complications in growth-restricted infants soon after birth [9].

Epidemiological studies have consistently reported that FGR fetuses present a higher risk of long-term health defects. These include neurodevelopmental impairment and cardiovascular or endocrine diseases in adulthood (e.g. obesity, type 2 diabetes mellitus, coronary artery disease, hypertension and metabolic disorders) [30], [31]. Figure 2.4 shows some of the most common physical and neurological problems observed in FGR infants.

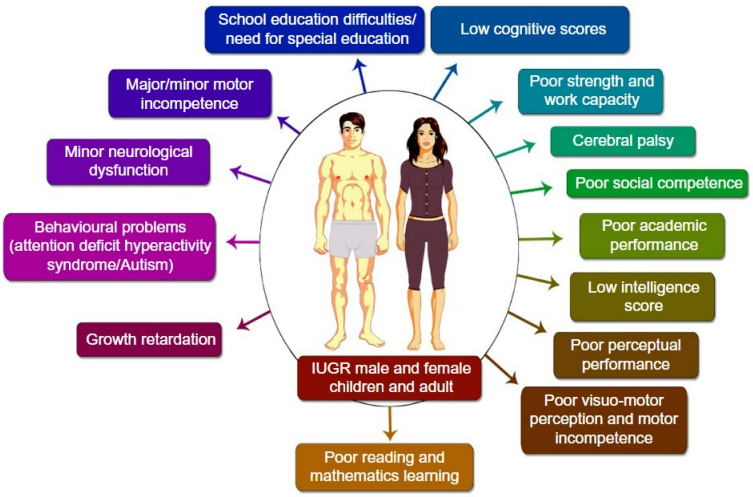

Figure 2.4: Physical and neurodevelopmental problems that remain in growth-restricted infants throughout time [9].

2.3 Ultrasound

US has been used for over half a century to scan the human body. The earliest recorded use of US for medical purposes was in the 1940s, when Dr. Karl Theo Dussik, an Austrian neurologist, employed it to image the brain [32], [33]. In the 1950s, the technique was also introduced in the field of Obstetrics and Gynaecology [33].

The widespread adoption of ultrasound can be attributed to several factors, such as its mobility, its noninvasive and radiation-free nature, its high temporal resolution, and its relatively low cost compared to other imaging modalities, such as magnetic resonance imaging and computed tomography [34], [35]. Consequently, it is a safe and real-time imaging technique that provides immediate visual guidance for numerous interventional procedures.

2.3.1 Characteristics of Sound Waves

A sound wave is a mechanical energy pressure wave transmitted through a medium by molecular vibration. The molecules in that medium oscillate back and forth, creating zones of compression (high pressure) and rarefaction (low pressure), as can be seen in Figure 2.5 [36]–[38].

Sound waves cannot travel in vacuum. They require a medium for transmission because, as previously stated, they are mechanical energy that is transmitted between molecules. They can be transmitted as longitudinal waves in gases or liquids (Figure 2.5), or as transverse waves in solid mediums (Figure 2.6).

In general, audible sound waves have frequencies ranging from 20 Hz to 20 kHz [35]–[37]. However, US waves are non-audible sound waves with a frequency that exceeds the range of the human hearing (above 20 kHz) [36]–[39]. In medical imaging, high-frequency waves are used (1 – 20 MHz) [38], [40].

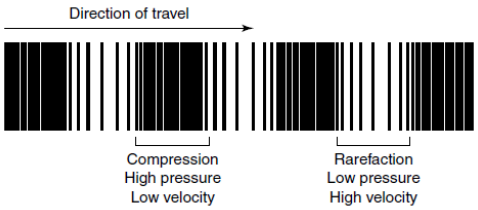

Figure 2.5: Longitudinal sound wave, where molecules oscillate in a linear direction [38].

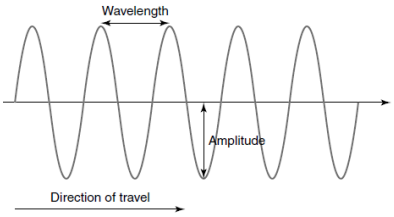

Figure 2.6: Transverse sound wave, where molecules oscillate perpendicular to the direction of the propagated wave [38].

2.3.2 Generation of an Ultrasound Wave

US transducers (also known as probes) contain multiple small piezoelectric crystals that generate US waves. When an electrical voltage is applied to the crystals, they convert electrical energy into mechanical vibrations through a phenomenon known as the reverse piezoelectric effect that will produce high-frequency sound waves [34], [36], [37].

Although each crystal produces a single US wave, the transducer sends many pulses through the human body. These pulses are timed and directed in a specific way to create a US beam that can penetrate tissues and create detailed images of structures like organs or blood vessels.

2.3.3 Generation of an Ultrasound Image

After the US beam passes through the body, it will reflect off the tissue interface and return to the transducer which acts simultaneously as a receiver [37]. Therefore, it will apply mechanical energy to piezoelectric crystals which will create an electrical signal that will be processed and presented on the screen as an image [36]. The ability to convert mechanical energy into electrical energy is known as the piezoelectric effect. There are different types of US acquisition modes: amplitude mode (or A-mode), brightness mode (or B-mode) and motion mode (or M-mode).

2.3.4 Tissue Interaction

The US beam interacts with tissues as it travels through the body, resulting in a loss of amplitude and energy that increases with the depth of penetration [36], [38]. This phenomenon, known as attenuation, is caused by a combination of absorption, reflection, scattering, and transmission (Figure 2.7).

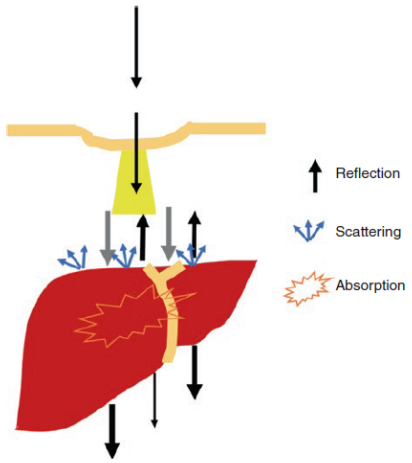

Figure 2.7: Reflection, scattering and absorption are effects of sound wave interaction with tissues that cause attenuation [36].

- **Reflection (or specular reflection):** when an US pulse hits a smooth interface between two tissues with different acoustic impedance, the energy is reflected back to the transducer. The amount of reflection depends on the degree of impedance mismatch between the tissues. This type of reflection happens at fascial sheaths, diaphragms, and the walls of major vessels [34], [36], [38].

- **Scattering (or diffuse reflection):** when an US pulse encounters an object smaller that its wavelength (e.g. red blood cells), or a rough, irregular body interface (e.g. surface of visceral organs), the energy is reflected in many directions, with only a fraction returning to the transducer at lower intensity [34], [36], [38].

- **Absorption:** as the US beam travels through a medium, its energy is absorbed, generating heat. This process is responsible for around 80% of the attenuation of the US beam in soft tissue [34], [36], [38].

- **Transmission (or refraction):** when an US pulse encounters an interface between tissues with different sound speeds, the direction of the beam changes. The degree of change depends on changes in velocity rather than acoustic impedance [36]. This process is one of the most common sources of artefacts in US image. Fat and bone are surfaces which cause a significant refraction [34], [36], [38].

2.3.4.1 Attenuation

Attenuation is represented by the attenuation coefficient (α) of the specific tissue and is directly proportional to the frequency of the US wave and inversely proportional to the distance travelled [36], [38]. Thus, high-frequency US waves are more attenuated (Figure 2.8), which means there is limited tissue penetration, and denser tissues generate higher attenuation rates.

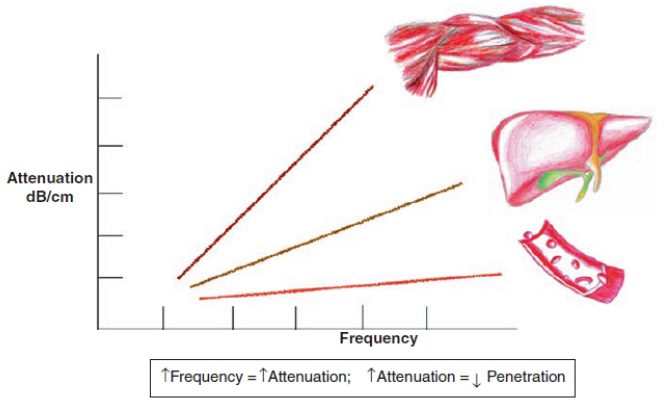

Figure 2.8: Relationship between attenuation and frequency in different tissues (muscle, liver and blood) [36].

2.3.4.2 Acoustic Impedance

Acoustic impedance is a tissue's resistance to the transmission of US and is a product of the density (ρ) of the medium and the acoustic velocity (c) of the US wave [36], [38]. As it is directly proportional to the degree of reflection, the higher the difference in acoustic impedance of tissues at the interface, the greater the amount of reflection [34].

Table 2.1: Acoustic impedance of various body tissues and organs.

Body Tissue	Acoustic impedance (10^6 Rayls)
Air	0.004
Lung	0.18
Fat	1.34
Kidney	1.63
Liver	1.65
Blood	1.65
Muscle	1.71
Bone	7.8

According to the acoustic impedance values listed in Table 2.1, the degree of reflection will be higher at air/soft tissues and bone/soft tissues interfaces since air has a very low acoustic impedance and bone has an extremely high acoustic impedance in contrast to other organs and tissues [34], [36], [38]. Because of the large acoustic impedance gradient, the resultant echoes are quite powerful in these instances. Conversely, low-intensity echoes are produced when two tissues have similar acoustic impedance values.

2.3.5 Tissue Echogenicity

Echogenicity of tissue refers to its capacity to reflect or transmit US waves in the context of the surrounding tissue. This will determine the degree of brightness of a dot on the image which

depends on the amplitude of the return echo. The final US image results from the combination of all the dots [36].

On the screen there will be a clear difference in contrast if there is an interface which exhibits different echogenicities [41]. So, according to their echogenicity (or brightness), structures can be classified as hyperechoic, hypoechoic or anechoic (Figure 2.9 and Figure 2.10). Table 2.2 shows echogenicity values of normal fetal tissues.

- **Hyperechoic:** areas that reflect echoes of higher energy than their surrounding structures, such as the diaphragm, gallstone, bone, and pericardium. These structures appear white on the screen [38].
- **Hypoechoic:** areas that reflect echoes of lower energy than their surrounding structures, such as solid organs. These structures appear grey on the screen [38].
- **Anechoic:** areas that do not reflect any echoes, such as fluid- and blood-filled structures. These structures appear black on the screen [38].

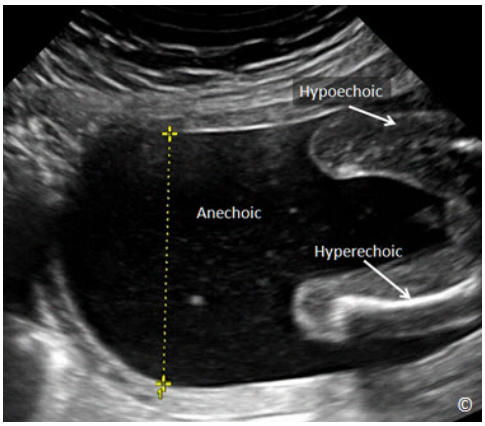

Figure 2.9: Second trimester US image of fetal extremities. Femur is hyperechoic compared to soft tissue, which is hypoechoic. Amniotic fluid appears anechoic in image [37].

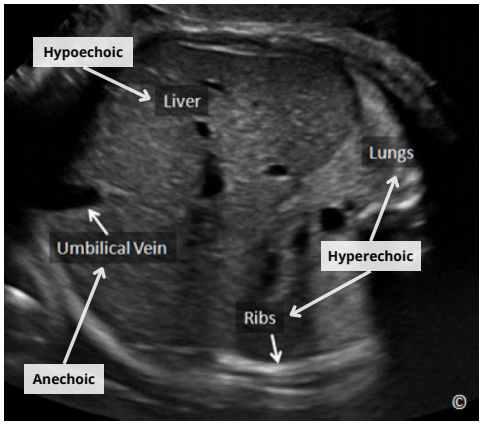

Figure 2.10: US image of fetal abdomen in the second trimester: ribs and lung tissue are hyperechoic, whereas liver is hypoechoic and umbilical vein is anechoic. Adapted from [37].

Table 2.2: Echogenecity of fetal tissues.

Organ System	Anechoic	Slightly Echoic	More Echoic	Echogenic
Bone	No	No	No	**Yes**
Brain	No	**Yes**	No	No
Lung	No	No	**Yes**	No
Stomach	**Yes**	No	No	No
Liver	No	**Yes**	No	No
Intestines	No	No	**Yes**	No
Kidney	No	No	**Yes**	No
Bladder	**Yes**	No	No	No
Placenta	No	No	**Yes**	No
Amniotic Fluid	**Yes**	No	No	No

2.3.6 Obstetric Ultrasound

Over the last three decades, the use of US has revolutionised medicine. In the 1980s, this type of medical exams became crucial for obstetricians and gynaecologists due to fast technology advances and real-time equipment introduction [42].

At the present, US is the most widely used medical imaging technique during pregnancy and it provides valuable information about fetal growth and development by assessing fetal biometry and organ functions [1]. It is useful in the detection of fetal growth anomalies and abnormalities (e.g. FGR) and it is a noninvasive medical imaging technique that does not use ionising radiation. So, it is safe for the mother, the fetus, or the person performing the exam.

The availability of high-quality obstetric US imaging is important for appropriate diagnosis and treatment. However, these images sometimes have a significant number of artefacts due to involuntary fetal movements at the early stage of pregnancy which can lead to possible misdiagnosis [1]. Additionally, this is a technique that presents a high degree of subjectivity in observing and obtaining standard measurements.

An obstetric US examination involves an assessment of fetal presentation and number, amniotic fluid volume, fetal cardiac activity, placental position, fetal biometry, and an anatomic survey [43]. There are two main types of fetal US imaging techniques (transabdominal US and transvaginal US) and they can be performed at different times throughout pregnancy. In the first trimester an US examination is done to confirm the presence of an intrauterine pregnancy, confirm fetal viability through the presence of fetal heart activity, diagnose or evaluate multiple gestation, evaluate a suspected ectopic pregnancy or miscarriage, estimate GA and assess basic fetal anatomy [33], [44]. Additionally, it can be used to screen fetal anomalies and aneuploidies in early pregnancy. In the second and third trimester an US examination is performed to evaluate fetal biometry, monitor fetal growth, estimate the amount of amniotic fluid, assess placental location and appearance, and observe fetal behaviour and activity [45].

2.3.6.1 Fetal Biometry

Fetal biometry, which is the US measurement of the anatomic segments of the fetus, has enhanced the ability to assess accurate GA, predict the day of delivery, detect growth abnormalities and diagnose fetal malformations and chromosomal syndromes [46].

It plays an important role during pregnancy and the most common biometric indices are CRL in first trimester of gestation and BPD, HC, AC and FL in second trimester of gestation.

- **Head Measurements**

  One of the most helpful and well-established metrics for determining GA is fetal head size. Although BPD was the first US parameter to be used for this purpose, studies have revealed that its accuracy could be affected by variations in fetal head shape (e.g. dolichocephaly and brachycephaly) and the position of the fetus in the uterus. Thus, HC has become the most widely used parameter to assess GA [47].

  - **Biparietal Diameter:** BPD is the largest axial dimension of the fetal skull and it is obtain from the proximal parietal bone's outer edge to the distal parietal bone's inner edge (outer to inner) (Figure 2.11-a), when cranial bones are perpendicular to the US beam. However, BPD can also be measured from the proximal parietal bone's outer edge to the distal parietal bone's outer edge (outer to outer) (Figure 2.11-b). The approach chosen varies on the country where the measurement is performed.
  - **Head Circumference:** HC is the outer perimeter of the skull and it can be either measured at the same level of BPD using the ellipse facility of US equipment or calculated using a geometrical formula (e.g. ellipse-traced).

- **Abdominal Circumference**

  Abdominal size is assessed by measuring the AC that is the outer perimeter of the skin line. Such as HC, it could be measured by using the ellipse facility of US equipment with the stomach and a segment of the umbilical vein at the level of the main portal vein [47].

  AC is used for estimation of fetal weight and it is the most sensitive parameter in predicting nutritional problems of the fetus as it is influenced by the thickness of abdominal wall and by the amount of hepatic glycogen reserves [46]. Although AC should not be used to determine GA, in isolation it is the most sensitive parameter for detecting FGR [48], [49].

- **Femur Length**

  FL is measured from the origin to the distal end of the shaft, from the greater trochanter to the lateral condyle. Despite not being included in the measurement of FL, both proximal and distal epiphyseal cartilage should be observed. Besides that, when the image is taken, the bone should be perpendicular to the US beam in order to see the entire extent of the femur on the screen and to avoid its foreshortening or elongation.

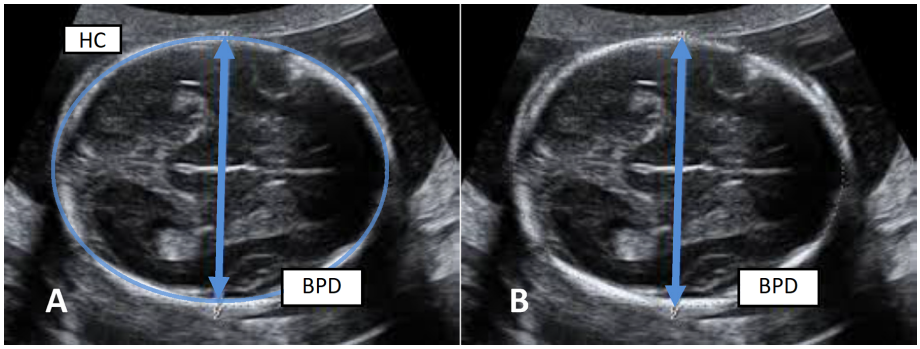

Figure 2.11: Fetal head biometric measurements: a) HC and BPD measured from outer to inner, b) BPD measured from outer to outer (B). Adapted from [47].

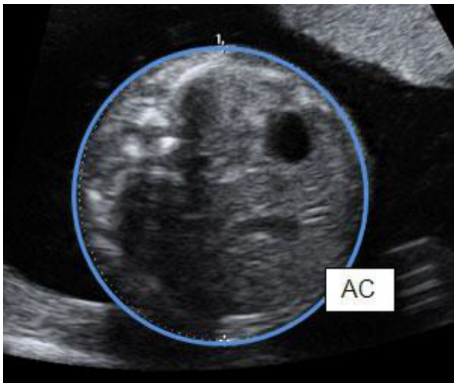

Figure 2.12: Fetal abdominal biometric measurement: AC [47].

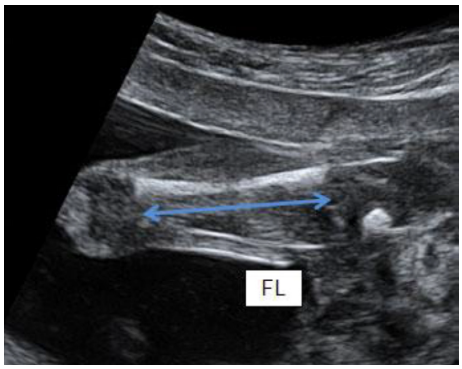

Figure 2.13: Fetal femur biometric measurement: FL [47].

2.4 Machine Learning

AI comprises all programs and algorithms that mimic human behaviour and replicate or outperform human decision-making [50]. ML approaches are in the branch of AI and have been widespread used in prominent sectors of application, including health care, forecasting, decision making, service performance improvement, automated tasks, and security. In the last few years, these techniques have proven to be quite useful in analysis of big data, image recognition, speech recognition, natural language processing and safety threat prevention [50], [51].

ML focuses on the development of algorithms that allow a computer to learn or enhance

its performance from past information available [52]. These algorithms use a built-in mathematical model that makes predictions without being explicitly programmed to perform a specific task [51]. In ML there are four types of learning algorithms that are used depending on the given problem and the available data: supervised, unsupervised, semi-supervised, and reinforcement learning. However, Supervised Learning (SL) and Unsupervised Learning (UL) are the most common.

2.4.1 Supervised Learning

In SL, there is known information regarding the target values or categories assigned to the instances used for training the model, i.e., the input data is labelled. The goal is to learn a mapping function and use it to predict the output for new unseen data [53].

SL models are generally used to solve two types of problems: classification, where the output is a discrete value or category, and regression, where the output value is a continuous value. Classification tasks are used when data can be categorised or separated into specific groups or classes. Therefore, in a binary classification there are always two possible outcomes, whereas in a multiclass classification there are more than two classes. Regression problems are similar to classification tasks. They may be formulated using a function approximation which tries to explain the relationship between two variables by fitting a linear equation to observed data [54].

Some of the most known SL models are Decision Tree (DT)s, Random Forest (RF)s, Naive Bayes (NB)s, Support Vector Machine (SVM)s, k-Nearest Neighbour (kNN)s, linear and logistic regressions, and Neural Network (NN)s.

2.4.2 Unsupervised Learning

In UL, the label information is not available. In these models, the goal is to analyse unlabelled data and detect hidden patterns in order to aggregate samples with similar behaviour [53], [54]. The most common UL tasks include clustering, dimensionality reduction and association rule learning [52].

In clustering, data is grouped into a certain number of subgroups, i.e., clusters, without having a predefined hypothesis on the features such clusters could have. On the other hand, dimensionality reduction consists in the transformation of each high-dimensional data point into two or more dimensions without affecting the data dispersion shape [51]. Association rule learning is a ML technique that uses rules to uncover interesting associations [52]. Some of the most known UL models are k-means clustering, density-based clustering algorithms, hierarchical clustering analysis and principal component analysis.

2.4.3 Machine Learning Pipeline

To develop a ML algorithm, the raw input data need to be analysed and processed since it may contain incomplete, inconsistent and noisy information. Data processing is important

because the performance of the model is related to the quality of the data used in the learning process.

Feature extraction and feature selection methods should be used to allow the creation of a dataset that emphasises the most relevant features for the model to learn. The resulting dataset is then subdivided into train and test set. The former is the set from which the model learns, whereas the latter is used to check the performance of model after training.

Before the learning process, it is important to choose the most suitable model to the task at hand. The train set is given as an input to the model which will find patterns and make predictions. After this training process, the model is tested on unseen data. The test set is then used to evaluate the performance of the model according to certain evaluation metrics. If the results obtained are reliable, the model is considered ready for implementation and evaluation in a real situation.

2.4.4 Evaluation Metrics

A learning model aims to learn a function that describes the observed data and is able to generalise when tested against unseen data. The evaluation metrics are used to estimate how well a given model performs in general and to choose the most suitable model throughout the comparison of the performance of different classifiers [53], [55].

For binary classification problems, there are always two possible values (positive and negative class) which represent the binary outputs (1,0) or (true, false) [51]. This categorisation into positive and negative classes allows the definition of four important concepts: True Positive (TP), True Negative (TN), False Positive (FP) (type 1 error) and False Negative (FN) (type 2 error). The number of positive and negative instances correctly predicted is called TP and TN, respectively. On the other hand, the misclassified instances, i.e., the number of positive and negative instances incorrectly predicted, are known as FP and FN, respectively.

Some of the most popular evaluation metrics will be presented according to the concepts mentioned above.

- **Confusion Matrix**

  In ML, the confusion matrix is a very useful tool that is used to represent classification models' performance. It is a square matrix with N dimensions, where N denotes the total number of target classes. This representation summarises the number of correct and incorrect predictions for each class and shows the distribution of the misclassified test samples. In Figure 2.14 it is possible to observe the confusion matrix for a binary classification problem.

- **Accuracy**

  The accuracy calculates the fraction of correctly predicted classes over the total number of instances of the dataset used for the evaluation. However, as this metric depends on the balance of the test data, it not always reflect the effectiveness of the learning model [54]. Equation 2.2 depicts the accuracy calculation for a binary classification problem.

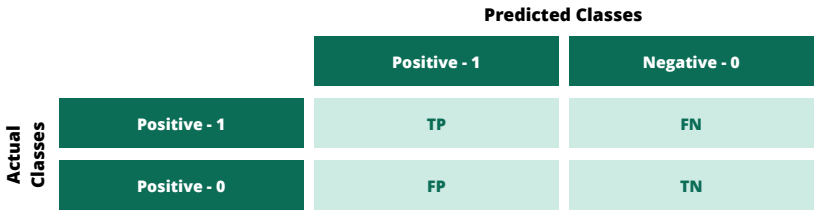

Figure 2.14: Confusion matrix of a binary classification problem.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{2.2}$$

- **Precision**

  The precision measures the fraction of predicted positive instances that are in fact real positive instances. Equation 2.3 depicts the precision calculation for a binary classification problem.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2.3}$$

- **Recall**

  The recall, also known as sensitivity, measures the fraction of real positive instances that are actually classified as positive (i.e. the True Positive Rate (TPR)). Equation 2.4 depicts the recall calculation for a binary classification problem.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{2.4}$$

- **F1-Score**

  The F1-Score represents the harmonic mean between precision and recall metrics and it must be used in order to overcome the accuracy measure's constraint in classification problems with imbalanced dataset [51]. Equation 2.5 depicts the F1-score calculation for a binary classification problem.

$$\text{F1-Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{2.5}$$

- **Specificity**

  The specificity measures the fraction of real negative instances that are actually classified as negative. Equation 2.4 depicts the specificity calculation for a binary classification problem.

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{2.6}$$

- **Area Under the Receiver Operating Characteristic (AUROC)**

  Sometimes, the result of a ML classifier can be expressed as a probabilistic score between 0 and 1. After setting a cut-off value, it is possible to define that: when the probabilistic score exceeds the threshold, the result is a positive case; otherwise, it is a negative case. The Receiver Operating Characteristic (ROC) curve is a probability curve that plots sensitivity (i.e. the TPR) against 1 − specificity (i.e. the False Positive Rate (FPR)) for different cut-off values. This curve depicts the evolution of correctly predicted positive cases against the increase in the number of false positives cases. Figure 2.15 shows the ROC curves for three different scenarios.

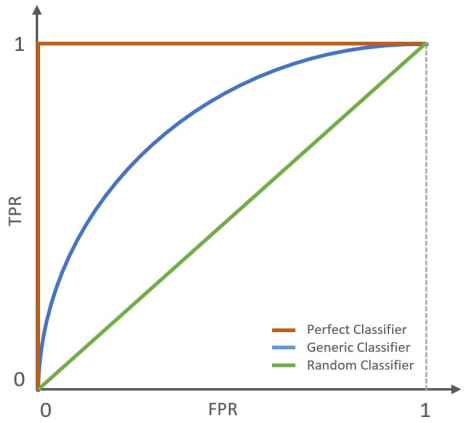

Figure 2.15: ROC curves for three different types of classifiers: perfect classifier, random classifier and generic classifier.

A perfect classifier is a model that can correctly distinguish between all positive and negative classes. So, the sensitivity and specificity of this model are both equal to 1 and its ROC curve has the behaviour described in the figure below. A random classifier is a model that has the same chances of producing a FP or a TP. Thus, its ROC curve is a diagonal line starting at the point $(0, 0)$ and ending at the point $(1, 1)$. Lastly, the ROC curve of a generic classifier generally lies between the curves of the previous models. The AUROC is a performance metric for evaluating classification problems. Thus, as can be seen in the above figure, a perfect model has an AUROC close to 1, whereas a random model has an AUROC close to 0.5. The former has a better ability to separate the different classes.

Chapter 3

Literature Review

This chapter provides a review of the current state of the art on the use of AI models in the field of obstetrics that contribute to breakthroughs in the diagnosis of FGR. As previously stated, fetal biometry gives the most accurate and important information regarding fetal growth and well-being. Although FGR is a fairly common clinical condition, its diagnosis remains a significant problem for obstetricians due to the difficulty in accurately defining fetal growth rate. Furthermore, US segmentation is very challenging because the quality of fetal US images is poor due to a low signal-to-noise ratio, inherent speckle and some artefacts like strong shadows produced by the spine or ribs in abdomen images, for example.

In this literature review, which will be subdivided into three sections, several methods will be presented. The first section focuses on the most commonly used approaches for FGR diagnosis in current clinical practice. In the second one, the evolution of computer-aided tools over time is explored and discussed. The last section gives an overview of both approaches.

3.1 Manual Methods

In clinical practice, FGR is diagnosed by identifying a fetus that is smaller than expected for its GA, either by physical examination or US [6]. The first one involves the measurement of SFH between 24 and 38 weeks of gestation, which is the distance between the upper border of the symphysis pubis bone and the top of the uterine fundus. Although this method is more simple and inexpensive, studies have shown that its accuracy in predicting fetuses with FGR is limited as it may be influenced by factors such as maternal obesity, uterine fibroids, amniotic fluid volume, fetal position and abdominal wall thickness.

The second method is the gold standard for detection of FGR. Standard fetal biometry includes assessment of AC, BPD, FL and HC, and fetal weight is estimated based on various

combinations of these four biometric indices using one of many published equations.

Warsof et al. [56] used the BPD and AC to estimate fetal weight for the first time. However, this method presented some limitations, namely a high prediction error such that the weight obtained was often below the expected value. In [57], Hadlock et al. proposed the use of two equations that combined various indices in the following way: the first equation should encompass the AC, HC and FL, and the second one should use the AC, BPD and HC. The results showed that this multi-parameter measurement method allows the estimation of fetal weight with a high level of accuracy.

Later, Hadlock [58] measured the AC, BPD and HC of 160 fetuses with macrocephaly and used two different equations to calculate their weight: one equation used the AC and BPD indices, and the other one used the AC and HC. Once again, the results obtained revealed that both equations underestimated the fetal weight. In another study, carried out in the same year, Hirata et al. [59], analysed 141 suspected macrosomia fetuses using three different equations. In this case, the results show that the equation that used the AC and FL indices was the one that presented the highest accuracy.

In recent years, several studies have focused on comparing the accuracy of different equations found in the literature. Most of them came to the conclusion that equations based on three or four biometric indices achieve more consistent outcomes. Milner and Arezina shown in [25] that the Hadlock equation (see equation 2.1 - section 2.2.3), which uses the indices AC, FL and HC, presents highest accuracy and is the most often employed in a clinical environment.

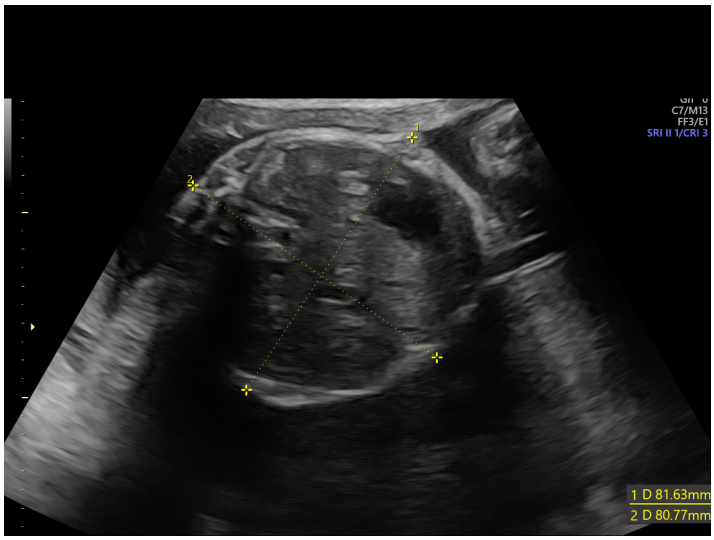

Figure 3.1: Manual delineation of the AC as the Region of Interest (ROI).

In current clinical practise, the biometric parameters described above should be gathered manually and in a standardised manner by clinicians because it is necessary for the correct interpretation of fetal biometry. Figure 3.1 shows a manual delineation of the AC on a US image.

However, the entire process is very challenging. Apart from manually tracing being time-consuming and user-dependent because it requires understanding of anatomy and spatial awareness, the intra- and inter-observer variability of clinicians influences the accuracy and consistency of measurements [60], [61]. Furthermore, some human limitations such as fatigue and repetitiveness of actions may affect the specialist's analysis (e.g. repetitive stress injury) and result in an inaccurate diagnosis [62]. For these reasons, the development and implementation of robust semi-automatic and automatic techniques for the extraction of fetal biometric measurements has recently gained attention in the hope of improving clinicians' workflow and overcoming operator dependence, increasing patient throughput, and improving measurement accuracy and consistency [62], [63].

3.2 Automatic and Semi-automatic Methods

3.2.1 Methods for AC segmentation

Over the last two decades, researchers have proposed a variety of semi-automatic and automatic algorithms for fetal biometry measurement. While many of these algorithms concentrate on measuring and segmenting fetal HC, some studies have also attempted to automate the evaluation of fetal AC.

In 2008, Yu et al. [64] proposed an automated method for detecting and measuring fetal abdominal contour using Iterative Randomized Hough Transform (IRHT). The proposed method is subdivided into four main steps that combine multiple image segmentation techniques. First, an enhanced instantaneous coefficient of variation edge detector is developed in order to identify edges of the abdomen contour and mitigate the impacts of most speckle noise. The Fuzzy C-Means clustering algorithm is then used to distinguish between the most prominent edges due to the abdominal contour and weak edges related to another texture. Therefore, the IRHT is applied to determine an elliptical contour and to obtain an initial AC estimate. Finally, a Gradient Vector Field (GVF) snake is used to adapt the initial ellipse to the actual contour of the fetal AC.

Also in 2008, Carneiro et al. [62] used a constrained probabilistic boosting tree classifier to segment the fetal head, abdomen and femur in US images, and to reproduce the respective standard biometric measurements. It is worth noting that this team has created and patented the first fetal biometry system that has been implemented in clinical practice. The system, called Auto OB, can detect all standard biometric measurements, as well as fetal humerus length and CRL. However, it needs larger fetal HC datasets.

In 2014, Wang et al. [65] presented an approach to detect and measure the AC from fetal US images. The proposed method is subdivided into two main steps: first, local phase information is used to detect the boundaries of fetal abdomen and second, an IRHT is employed in order to detect the ellipse that fits to the abdominal contour. The methodology was tested in 90 fetal US images, and experimental findings show a high level of agreement between manual and the proposed method for measuring AC (mean sign difference is 0.42% and correlation

coefficient is 0.9973).

The acquisition of a standard plan is a prerequisite for an US diagnosis. Since this is a crucial step for subsequent biometric measurements and diagnostics, in 2015, Chen et al. [66] proposed a learning-based approach to automatically locate the Fetal Abdominal Standard Plane (FASP). To do this, a domain transferred deep Convolutional Neural Network (CNN) was employed.

In 2016, Ravishankar et al. [67] proposed a hybrid approach for the automatic detection and measurement of AC from Two-dimensional (2D) fetal US images. The authors tested their work on 70 images and achieved a mean dice similarity coefficient of 0.90. Two years later, Kim et al. [68] proposed an automatic AC measurement approach based on multiple combinations of CNNs and a U-Net. This method is divided into three main steps: the first one is related to the estimation of the initial AC, the second one concerns the AC measurement, and the last one corresponds to plane acceptance checking. The method was tested on clinical US data acquired from 77 pregnant women and the results showed a dice similarity coefficient of $(92.55 \pm 9.1)\%$ and an accuracy of 87.10% for the acceptance check of the FASP.

In 2020, Yang et al. [69] developed an algorithm capable of automatically segmenting fetal anatomical structures in 2D US images. A RF algorithm was used to transfer the acquired image to a primary classification map that would allow the recognition of the ROI anatomical structures. In order to obtain a more accurate classification map, a Scale-aware Auto-context Model (SACM) model was used, whose goal was to improve the map contour details from several visual levels. The results revealed that the model demonstrated a good ability to segment both AC and HC parameters.

In 2021, Gao et al. [70] proposed a NN for fetal weight prediction. The aim of this algorithm was to compare the estimation made by the network with values obtained through the equations mentioned above. A total of 80 cases of pregnant women were studied and the results showed that this approach allowed reducing the error in the prediction of fetal weight, but the accuracy of the estimate for the extremes of the weight range considered was not satisfactory.

3.2.2 Methods for FGR diagnosis

Regarding the diagnosis of FGR, some studies have been performed whose aim is to distinguish FGR from normal fetuses. In 2009, Ferrario et al. [71] applied a multiparametric classifier based on k-mean cluster analysis to identify FGR fetuses. This study was based on the analysis of fetal heart rate variability (FHR) performed in the antepartum period. The results achieved an accuracy of up to 82.4%.

Also in 2009, Nithya and Madheswaran [72] presented an automatic method where fetal AC extraction is required to identify FGR fetus. First, a speckle reducing anisotropic diffusion filter is applied to remove the speckling artefacts present in US images and IRHT is used to find the initial contour of AC. Then, to detect the outer edge of the abdominal contour, a GVF snake is used. The results showed that fetal growth can be easily assessed without being overestimated.

In 2014, Gadagkar and Shreedhara [73] developed an Artificial Neural Network (ANN) classifier to diagnose and classify FGR fetuses based on features from US images. The classification accuracy was greater than 90%. On the other hand, that year, Bagi and Shreedhara [74] proposed an automated method to assess fetal biometrics from 2D US images, and diagnose FGR. Two ANN models were used: multilayer perceptron (MLP) using back propagation algorithm and radial basis function (RBF).

In 2016, Wosiak et al. [75] performed a study that was more focused on the identification of FGR types. By comparing some SL and UL algorithms, the results revealed that hybrid classification methods show a better ability to distinguish the symmetrical from asymmetrical type of FGR, and UL models are not the most suitable for classification.

3.3 Overview

Several methods for automating the segmentation and measurement of fetal AC were discussed throughout this chapter. Methods for detecting growth patterns and abnormalities, such as FGR, were also presented. Table 3.1 summarises the literature previously reviewed.

Automated analysis of US images is challenging, and methods established for other imaging diagnostic tools not necessarily work in these images. Furthermore, there are no standard methods for segmenting US images, and the segmentation strategies are application dependent. In the fetal imaging domain, previously published automated segmentation methods focused on using segmentation as an intermediary processing stage to estimate standard biometric measurements. Most of them were based on morphological operators, active contour models, Hough transform, deformable models or ML approaches.

To date, much of the research in fetal biometry has focused on identifying and segmenting the fetal head, as segmentation of the fetal abdomen is more challenging. The fetal head seems to be the easiest biometric measurement to identify and segment, as the boundaries set by the skull bones are precise and there are texture similarities between individuals. On the other hand, segmentation of the abdomen is more challenging because the abdomen does not have clear boundaries and its internal structures differ from individual to individual [62].

Some FGR investigations were also presented. However, automatic or semi-automatic detection of FGR patterns is still an open topic. Most studies attempt to distinguish FGR from normal fetuses, but they do not distinguish between the various types of FGR.

Overall, the algorithms presented throughout this chapter have certain advantages over the manual methods currently used in clinical practise. Aside from a large reduction in execution time, as they become less and less user-dependent, these approaches have demonstrated a good performance in terms of segmentation, producing results that are quite similar to those produced using manual methods. However, there are still two primary limitations to these algorithms: the need for a large and diverse database, as well as the high computational costs involved. Furthermore, the error rates of algorithms remain a critical issue that sometimes limits their implementation in clinical contexts.

Table 3.1: Overview of literature studies on the use of ML in Obstetrics.

Reference	Model Aim	Methodology
Yu et al. [64]	Automatic extraction and measurement of fetal abdominal contour	Fuzzy C-Means clustering algorithm, IRHT and GVF snake
Carneiro et al. [62]	Automatic detection and measurement of fetal anatomies	Constrained probabilistic boosting tree classifier
Ferrario et al. [71]	Identification of FGR fetuses	Multiparametric classifier based on k-mean cluster analysis
Nithya and Madheswaran [72]	Detection of FGR fetuses using fetal AC	IRHT and GVF snake
Wang et al. [65]	Detection and measurement of abdominal contour	Local phase-based method and IRHT
Gadagkar and Shreedhara [73]	Diagnosis and classification of FGR fetuses in 2D US images	ANN classifier
Bagi and Shreedhara [74]	Biometric measurement and FGR classification	Two ANN models: multilayer perceptron using back propagation algorithm, and radial basis function
Chen et al. [66]	Automatic localisation of the FASP	Domain transferred deep NN
Ravishankar et al. [67]	Automatic segmentation of fetal abdominal contour in 2D US images	Hybrid approach combining CNNs and gradient boosting machine
Wosiak et al. [75]	Automatic identification of FGR types	Hybrid classifiers algorithms, single classification methods and clustering
Kim et al. [68]	Automatic fetal AC estimation from 2D US images	Multiple combinations of CNNs and a U-Net
Yang et al. [69]	Automatic segmentation of fetal anatomical structures in 2D US images	Structured RFs and a SACM
Gao et al. [70]	Fetal weight prediction	Back propagation NN optimised by genetic algorithm

The next chapters will explore the proposal of this thesis: a semi-automatic approach capable of performing segmentation and textural analysis of the fetal liver in order to establish a possible correlation between hepatic echogenicity and the diagnosis of FGR.

Chapter 4

Materials and Methods

4.1 Materials

The data used in this study is from the database of the Department of Gynaecology and Obstetrics of the HGO which contains hundreds of US images. These images were acquired from pregnant women who received a routine US screening exam between January 2021 and July 2021.

As the proposed study focuses on the analysis of fetal liver echogenecity, only images of abdominal region were considered. Therefore, a total of 43 two-dimensional (2D) US images for different fetuses were effectively used: 23 cases of normal growth and 20 cases of FGR. The gestational age of the fetuses taken for analysis are between 30 to 37 weeks.

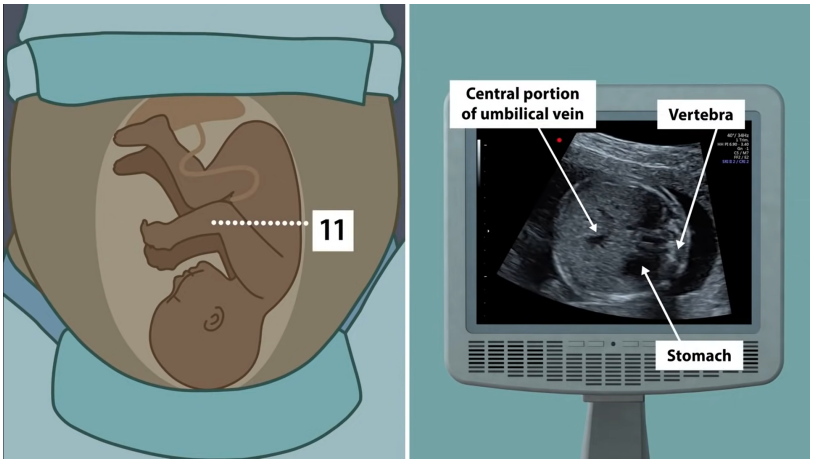

Figure 4.1: Transverse plane of the fetal abdomen. During an US examination, plane 11 is the most suitable for assessing fetal AC. From [76].

During image acquisition, a experienced sonographer examined the whole abdominal area and saved a single frame that was thought to be the most suitable for annotation of the fetal AC. For this purpose, it may have been required to move the probe in circular and angular movements in order to maintain a transverse orientation of the fetus. According to twenty planes and two sweeps approach of the International Society of Ultrasound in Obstetrics and Gynecology (ISUOG) [1], where each plane relates to a specific fetal view and has a number of structures to be evaluated, plane 11 (transverse plane) is where the landmarks of the fetal abdomen are visible (Figure 4.1). The AC measurement is then performed manually by placing the ellipse caliper on the outer edge of the fetal abdomen.

All images were acquired using the Voluson E10 US device (GE Healthcare) and each was stored in Digital Imaging and Communication in Medicine (DICOM) format with a spatial resolution of 1136×852 pixels. Additionally, they are anonymous and do not contain any personal information, including the medical file identification number and any other direct or indirect identifiers of the patients. The HGO ethics committee approved the use of these data for this study and informed consent was waived.

4.2 Methodology

This dissertation is an observational, cross-sectional, and retrospective study that aims to evaluate the following hypothesis: the liver of fetuses with and without FGR does not differ in terms of echogenicity. To that purpose, an AI model was developed that examines the texture of the fetal liver on US images before automatically classifying them.

The first step involved the analysis of images from the HGO database in order to ensure that only images of the fetal abdominal area were used in the course of the study. Next, the images were subjected to a ROI selection step to isolate the fetal abdomen. For this purpose, the yellow marks produced by placing the ellipse caliper on the outer edge of the abdomen during image acquisition were consider. This allowed the creation of a ROI that would be used to aid the segmentation of the fetal liver.

Initially, it was thought to carry out this method using a similar approach to that was used in [72]. However, the results of the initial abdominal contour acquired by IRHT and the final contour produced by GVF snake were not satisfactory: the percentage of cases that were incorrectly identified was higher than the percentage of cases that were correctly classified. As a result, this approach was discarded, and morphological operations (erosion, dilatation, opening and closing), connected components algorithm and contour detection method based on image moments were used to delimit the fetal abdomen.

In a second phase, a computer-aided tool was developed to enable the manual selection of a portion of the fetal liver. It was still considered to make this procedure automatic and to include the manual selection option solely in circumstances when clinicians needed to readjust the contour. However, as previously stated, it is challenging to detect tissue boundaries

[1] As stated in [45], ISUOG "is a scientific organisation that encourages safe clinical practice and high-quality teaching and research related to diagnostic imaging in women's healthcare".

in images with significant artefacts (like fetal USs) due to noise interference. It would be extremely difficult to automatically identify the liver because images of the fetal abdomen lack well-defined boundaries and contain internal structural inconsistencies.

After that, feature analysis was carried out. Taking inspiration from some works that pointed out the benefits of evaluating the textural features of the liver in US images [77], [78], two statistical methods of texture analysis were used in this study to extract features from each liver segment. After features extraction, a selection of the most relevant features for the study was proceeded. This process is extremely important because, by removing the most redundant and irrelevant features, it is possible to improve the predictive ability of the algorithms and reduce the overfitting.

The subsequent step consisted of developing the ML model through the implementation of the following supervised ML models: DT classifier, a RF classifier, a SVM classifier, a kNN classifier and a NB classifier. Then, a comparative evaluation was performed to choose which model best suited the proposed task. The whole processing pipeline is represented in Figure 4.2 and the illustrated steps are further described in the following sections.

The proposed approach was developed using Python programming language and the interface was implemented using PyQt.

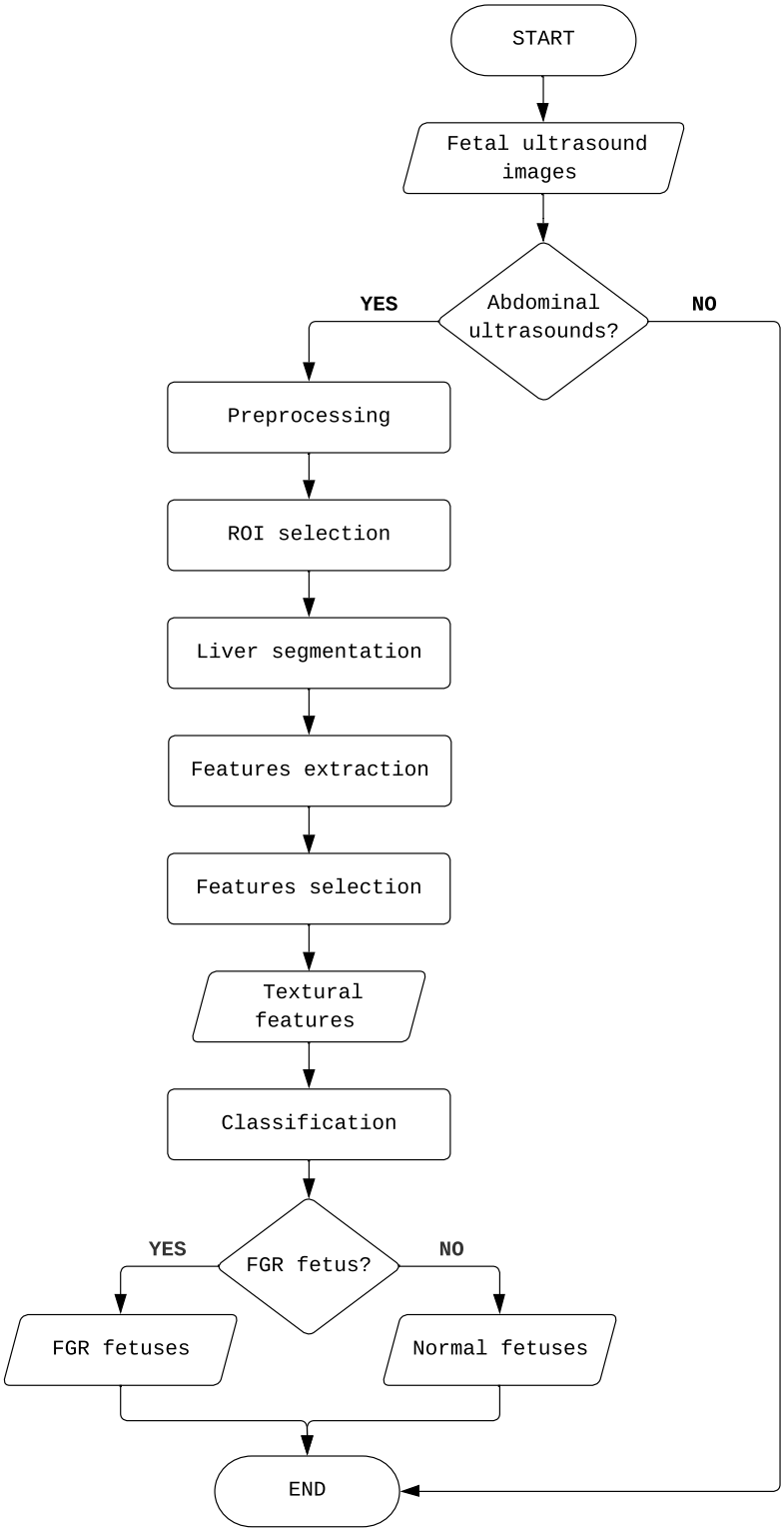

Figure 4.2: Flowchart of the proposed methodology.

4.2.1 Image Preprocessing

The HGO database contains hundreds of US images that were acquired during routine examinations performed on pregnant women. However, as mentioned earlier, only a very particular set of images was considered for this study.

Thus, this first step is required to create a more restricted and representative database. To simplify this selection process, Zegami platform was used. It is an AI-enabled image analysis tool which finds patterns and trends in large data sets of images and groups them into clusters.

Since the liver is part of internal structure of the fetal abdomen, the next step aimed to find a way to isolate the abdominal area. However, first and foremost, a binarization method had to be applied. The end result of this preprocessing was a set of binary images that match exactly the same number of original images.

4.2.1.1 Binarization Method

Binary images are a common type of image used in image processing and computer vision which have only two intensity values. They are created by thresholding a greyscale or colour image, where the pixels are assigned either a value of 0 for black or a value of 1 (or 255) for white [79].

The thresholding process involves setting a threshold value, and then assigning a pixel a value of 1 if its intensity exceeds the threshold value. Otherwise, pixels that do not exceed the threshold value are assigned a value of 0. This process typically produces a binary image with a white object on a black background. However, depending on the image to be thresholded, this polarity can be reversed.

Since the images used in this study include a very distinct RGB component, the binarization was performed as follows: pixels with different R, G and B channels were set to 1, while the remainder were set to 0. Hence, it was possible to highlight the four yellow marks, hereafter be known as calipers, that would later be used to segment the fetal abdomen. Figure 4.3 shows an example of the original and binary images.

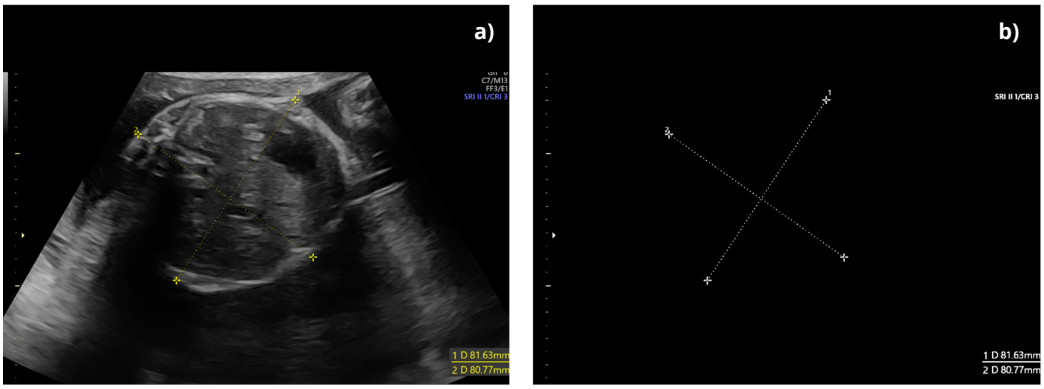

Figure 4.3: Binarization method. Original image a) and the resulting binary image b).

4.2.2 ROI Selection

The liver is an organ with a non-linear anatomy which is located in the abdominal region of the fetus. Its identification and measurement might be difficult because fetal structures are almost always occluded at the second and last trimester of pregnancy. Besides that, as already mentioned, fetal US images contain some artefacts and an inherent speckle noise.

In the ROI selection step, the strategy adopted was based on isolating the fetal abdomen using basic image processing techniques. Because the automatic segmentation of the liver proved extremely difficult, this step attempted to obtain a more restricted ROI to identify the organ under research more easily. The fours calipers were used to automatically obtain the outline of the abdomen and limit the image to its interior. A series of morphological operations and a connected component analysis were applied in order to identify these four points of interest. Therefore, an interpolation was made to remove the area outside the contour of the abdomen.

4.2.2.1 Morphological Operations

Morphological operations are image processing techniques which deal with the shape or morphology of features in an input image. The most basic operations are dilation and erosion which can be customised by the proper selection of a Structuring Element (SE). This element, also called as kernel, can be a matrix of odd size (e.g. 3, 5, 7) and defines how the objects will be processed [79], [80].

In dilation, the rationale is to increase the object area by filling small holes and connecting disjoint objects. In this case, each pixel is replaced by the maximum among its neighbours that are covered by the SE, i.e. each pixel in the original image is 1 if at least one pixel of the SE is 1. On the other hand, erosion is an operation of shrinking or thinning objects. In this case, each pixel in the original image will only be considered 1 if all pixels under the SE are 1, otherwise, it will be eroded (i.e. takes the value 0). Hence, in dilation the white region in the image increases while in erosion the size of the foreground object decreases [79].

As can be seen in Figure 4.3, there are several points surrounding the calipers. Due to their proximity, they need to be removed in order to isolate and obtain the centre of each caliper. Furthermore, in the images there is a lot of information that must be removed such as the text information in the top and bottom right corners. Thus, five iterations of a dilation operation were applied to the binary image. In order to do that, a 5-dimensional matrix was used as a SE. The kernel runs through the image five times, replacing to white all pixels patches where the kernel partially or fully fits. The result of this operation is shown in Figure 4.4-a. Next, a connected component analysis was performed to achieve the result shown in Figure 4.4-b.

4.2.2.2 Connected component analysis

Connected component analysis is a widely used algorithm in computer vision to detect objects in images that have irregular shapes and sizes, taking into account the connectivity between pixels. The algorithm is based on two types of connectivity: 4-connectivity and 8-connectivity. In 4-connectivity, pixels are connected to their neighbours only horizontally and vertically, while in 8-connectivity, pixels are connected horizontally, vertically, and diagonally [81].

The algorithm typically scans the image pixel by pixel, examining each pixel's connectivity to its neighbouring pixels. Pixels that are found to be connected based on a specified connectivity criteria are assigned the same label, creating a distinct object (or component) in the image. After the labeling step, the input image is transferred to a labeled image, where each object is represented by a unique label. This labeled image enables easy extraction of each object based on its label and calculate its shape features, such as area, major and minor axis length, perimeter, centroid, extent, orientation, and solidity. However, the use of this algorithm often requires a prior segmentation procedure using morphological operations.

As mentioned in the previous section, besides the contour of the central region, other contours were identified in the current image (Figure 4.4-a). To isolate the largest contour, a connected component analysis was applied with a 4-pixel adjacency. This allowed the selection of the component that corresponds to the four main points (Figure 4.4-b).

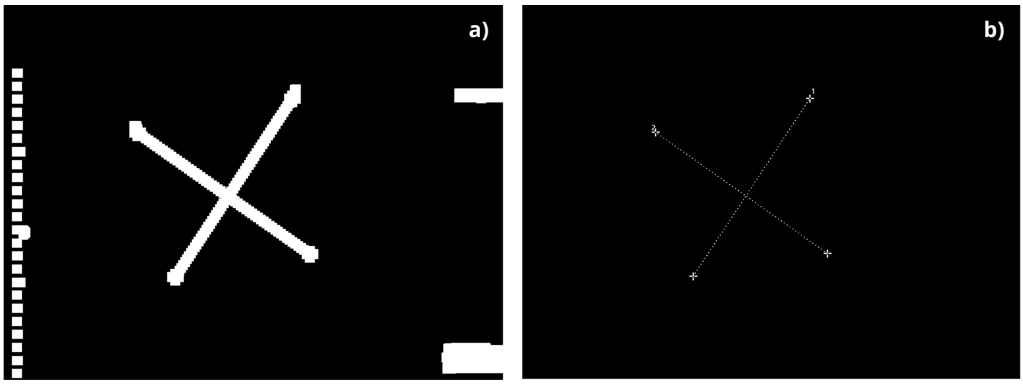

Figure 4.4: Morphological operations and connected component analysis applied in the processing of the binary image: (a) dilated image, (b) resulting image after a connected component analysis performed to the image in (a).

However, the outcome still does not allow obtaining the contour of the whole fetal abdomen. As a result, morphological operations had to be used once more to obtain the result of Figure 4.5. In this case, an opening operation was applied which is equivalent to applying a dilation right after an erosion. During the erosion operation (Figure 4.5-a), the SE scans the image and converts to black all the pixels patches where the kernel does not completely fits. During the dilation operation (Figure 4.5-b), SE scans the image and converts to white all the pixels patches where the kernel partially or completely fits. This time, two distinct SEs were employed. In the first situation, a 3-dimensional matrix was used, whereas in the second, a 7-dimensional matrix was used.

Finally, it was necessary to find the centroid of each caliper by calculating the moments of Figure 4.5-b.

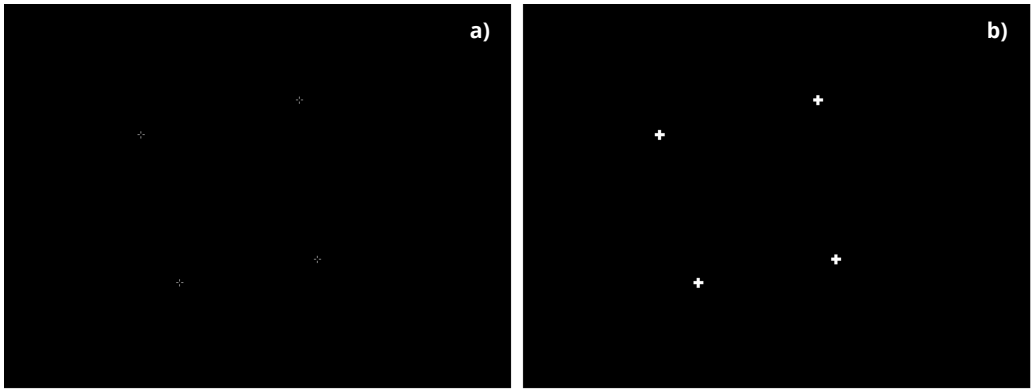

Figure 4.5: Opening operation result: (a) eroded image, (b) dilated image. Note that the dilation operation was performed to the image in (a).

4.2.2.3 Image Moments

An image moment is very useful in object recognition and shape analysis. It is a particular weighted average of pixel's intensities which allows the study of some simple image properties such as the area and centroid [82].

The image moment M_{ij} of order (i + j) for a greyscale image with pixel intensities I(x,y) is defined as:

$$M_{ij} = \sum_{x} \sum_{y} x^i y^j I(x, y) \tag{4.1}$$

For a binary image, the 0^{th} order moment corresponds to the object's area because it corresponds to the total number of non-zero pixels. So, using the Equation 4.1, the area of a given object can be obtained as follows:

$$M_{00} = \sum_{x} \sum_{y} I(x, y) \tag{4.2}$$

On the other hand, the centroid of an object corresponds to the arithmetic mean position of all points and is given by the following relation:

$$Centroid\{\overline{x}, \overline{y}\} = \left\{ \frac{M_{10}}{M_{00}}, \frac{M_{01}}{M_{00}} \right\} \tag{4.3}$$

Here, for a binary image, M_{10} corresponds to the sum of all non-zero pixels along the x-axis, while M_{01} corresponds to the sum of all non-zero pixels along the y-axis. M_{00} was mentioned above.

As shown in Figure 4.6, from the centroids of the four calipers (Figure 4.6-a) it was possible to trace the curve that best fits the approximate shape of the fetal abdomen (Figure 4.6-b). First, to identify the centroids it was necessary to calculate the moments of each object and find the x and y coordinates of each centroid using Equation 4.3. Following that, spline interpolation

was used to outline the fetal abdomen. For this purpose, a spline representation of a curve in a 2-D plane passing through the coordinates of the four centroids was required.

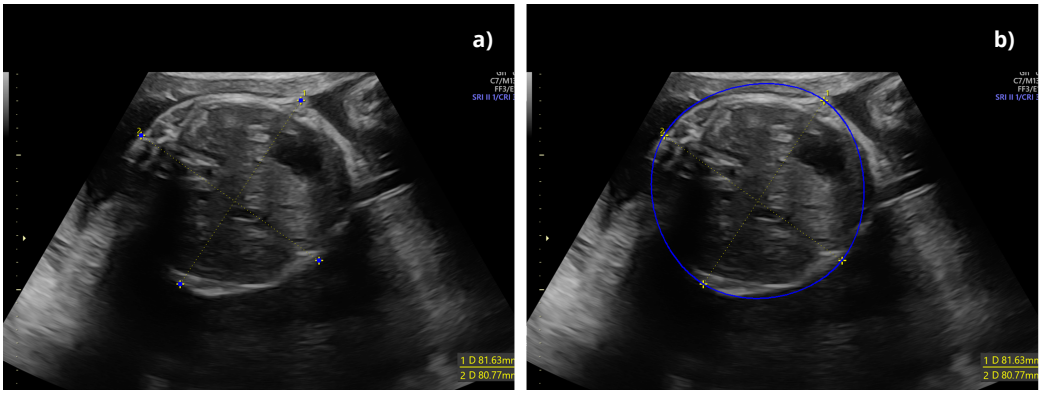

Figure 4.6: Curve fitting: a) centroids of the four main points, b) curve that best fits the approximate shape of the fetal abdomen.

After curve fitting, image masking had to be used to create a ROI. As the fetal liver is segmented only inside this region, the ROI mask and the corresponding original image, both with the same resolution, were combined using a bitwise operation. The resultant image is shown in Figure 4.7 where only the pixels outside the specified ROI have been modified and set to 0.

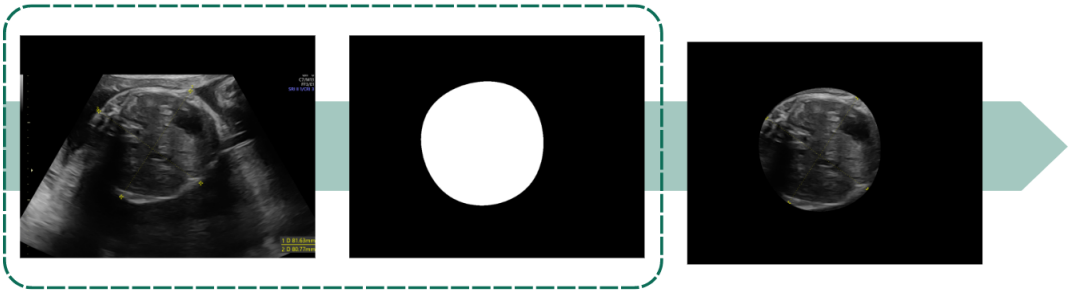

Figure 4.7: Fetal abdominal US segmentation outcome after applying a ROI mask to the original image.

4.2.3 Liver Segmentation

In this section, the process behind the manual delineation of the fetal liver on US images is explained in more detail. For this purpose, Bézier curves were employed to link the points defined along the fetal liver. These curves are intended to approximate, with a minimum of distortion, real world shapes that have no mathematical representation or whose representation is unknown.

A very user-friendly tool was developed to make this task more intuitive. It was design to be used by clinical professionals with the aim of segmenting a part of the fetal liver. Despite

being functional, this tool has yet to be evaluated in a clinical context. Its functionalities allow it to fulfil the minimum requirements, but it is still subject to a process of improvement. A more detailed analysis of this tool can be found in Chapter 5.1.

4.2.3.1 Bézier Curves

A Bézier curve is a series of control points that range from P_0 to P_n. The curve's order is determined by n, whereas the initial and final control points are always the curve's endpoints. As regards the intermediate control points, if any, they do not necessarily lie on the curve [83]. So, a Bézier curve of degree n is defined as

$$B(t) = \sum_{i=0}^{n} b_i^n(t) P_i, \quad 0 \leq t \leq 1 \tag{4.4}$$

where the coefficients P_i are the control points (or Bézier points) and the polynomials b_i^n defined as

$$b_i^n = \binom{n}{i} t^i (1-t)^{n-i}, \quad i = 0, ..., n \tag{4.5}$$

are the Bernstein basis polynomials of degree n.

- **Linear Bézier curve (n = 2)**: a curve that has two control points, P_0 and P_1. The curve is given by

  $$Q(t) = (1-t)V_0 + tV_1, \quad 0 \leq t \leq 1 \tag{4.6}$$

  and is equivalent to a linear interpolation.

- **Quadratic Bézier curve (n = 3)**: a curve that has three control points, P_0, P_1, and P_2. The curve is given by

  $$Q(t) = V_1 + (1-t)^2(V_0 - V_1) + t^2(V_2 - V_1), \quad 0 \leq t \leq 1 \tag{4.7}$$

  and is equivalent to a point-to-point linear interpolation of two linear Bézier curves.

- **Cubic Bézier curve (n = 4)**: a curve that has four control points, P_0, P_1, P_2 and P_3. The curve is given by

  $$Q(t) = (1-t)^3 V_0 + 3(1-t)^2 t V_1 + 3(1-t)t^2 V_2 + t^3 V_3 \tag{4.8}$$

  and is equivalent to a point-to-point linear interpolation of two quadratic Bézier curves.

During fetal liver delineation, the points set along the liver segment are connected by cubic Bézier curves in order to allow a smooth delineation that easily suits to the organ under study. Figure 4.8 shows that this curve fitting requires obtaining a tangent segment to the curve at each inserted point. These segments and their extremities (also known as intermediate control points) may be produced using the following methodology:

1. Consider previous and next points (e.g. A and C);

2. Find the angle bisector between the segments created with the previous and current point (e.g. $\overline{AB}$), and between the current and next point (e.g. $\overline{BC}$);

3. Create two segments perpendicular to this angle that start from the current point and have a proportional length for each segment;

4. Use the segments' extremities as control points.

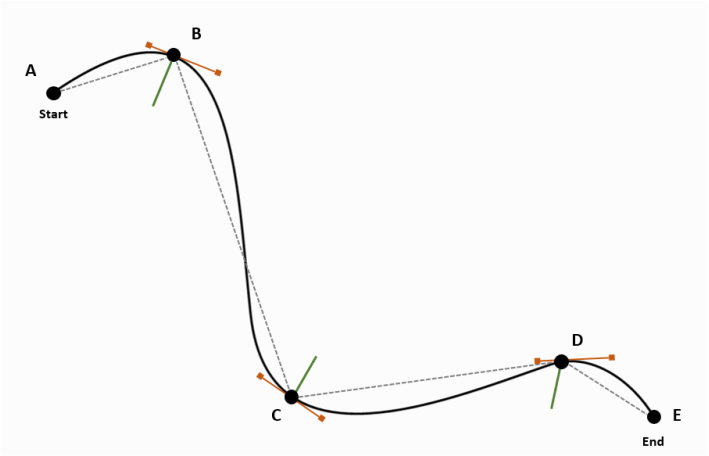

Figure 4.8: Spline interpolation using cubic Bézier curves.

Finally, it should be noted that the first and last curves are the only ones that are not cubic Bézier curves. In these two cases, quadratic Bézier curves were used since there is only one control point: the target line reference for the first point and the source line reference for the last point, respectively.

4.2.4 Feature Analysis

The ongoing study in the field of automatic identification of changes in fetal liver echogenicity aims to extract representative features in order to obtain a more reliable diagnostic method. Therefore, given the different texture types present in the segmented images, a texture analysis approach was employed to extract textural features from each segment in the feature extraction stage.

Texture analysis is a technique that was developed as a tool for image analysis and classification. In medical imaging, it became widely employed for the examination of US images of the liver and heart in the late 1970s and early 1980s. Later, it became quite popular for a wide range of medical imaging applications [84]. Texture analysis involves analyzing variations in pixel intensity or their spatial distribution to uncover patterns in images. It can be assessed using several approaches, including structural, statistical, model-based, and transform-based methods. Statistical analysis methods, which are commonly used in medical applications, use

the grey-level distribution within an image to describe its texture. These methods can be classified as first-, second- and higher-order statistics based on the number of pixels defining the local feature.

According to the information present in Table 4.1, only the first- and second-order statistics methods were used in this work.

4.2.4.1 First-Order Statistical Features

First-order statistical features, also called histogram features, provide information about the different statistical properties that are derived from the image histogram. These features provide information about the probability of a particular grey level occurring within an image, but do not take into account the spatial relationships or correlations between image pixels [85], [86]. The following histogram-based features were extracted from each liver segment:

- **Mean** is the average of the bright and dark values in the image.
- **Variance** is a measure of image contrast.
- **Skewness** is a measure of the degree of asymmetry.
- **Kurtosis** is a measure of the peak and flatness in relation to a normal distribution.

Their mathematical expressions are explicit in Table A.1 of Appendix A.

4.2.4.2 Second-Order Statistical Features

Second-order statistical features, also known as co-occurrence matrix features, describe the statistical properties of two pixels based on the Grey Level Co-occurrence Matrix (GLCM), a powerful tool used in image processing to capture the spatial relationship between pixels in an image. As a result, unlike first-order statistical features, interaction between image pixels is taking into account [85], [86].

In GLCM approach, knowing the total number of grey levels in a given image, it is possible to determine the probability of occurrence of a grey level i in the neighbour of another grey level j at a given distance d and direction θ. So, GLCM is a tabular representation of probabilities [87].

For any matrix M, GLCM matrix G is defined as:

$$G(x,y) = \sum_i \sum_j \begin{cases} 1, & \text{if } M(i,j) = x \text{ and } M(i+\Delta i, j+\Delta j) = y \\ 0, & \text{otherwise} \end{cases} \tag{4.9}$$

where, Δi and Δj represent the offset in row and column directions, respectively. Figure 4.9 shows four examples of GLCM matrix formation for a given matrix M, along four different directions. It can be observed that each direction is represented by a certain colour: 0° by red

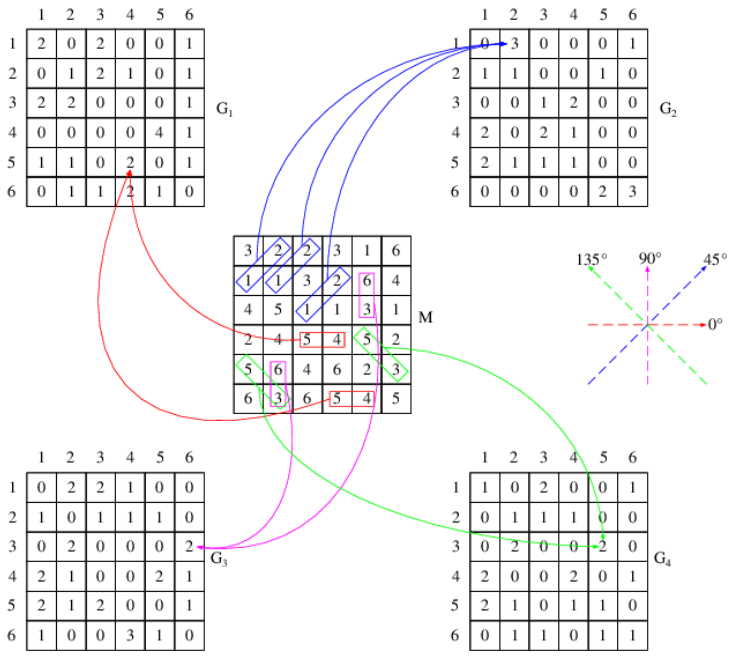

Figure 4.9: GLCM matrix formation. It can be seen that $G_2(1,2) = 3$, which indicates that in matrix M, the values 1 and 2 appear in pairs twice in 45° direction. The same rationale applies to the remaining GLCM matrices [88].

colour, 45° by blue colour, 90° by pink colour and 135° by green colour. After GLCM matrices formation, there are several features that can be calculated from each of them.

In this study four distinct co-occurrence matrices were computed for each liver segment: first had a horizontal offset of five ($\theta = [0]$ and $d = [5]$), second had a diagonal offset of five ($\theta = [45]$ and $d = [5]$), third had a vertical offset of five ($\theta = [90]$ and $d = [5]$), and last one had a anti-diagonal offset of five ($\theta = [135]$ and $d = [5]$). Then, it was necessary to normalise them, so they could be understood as probabilities of pairs of values occurring in a given image, and compute the following GLCM features:

- **Angular Second Moment (ASM)** measures the uniformity of the distribution of grey levels in the image. A higher ASM value indicates a more uniform distribution of grey levels.

- **Contrast** quantifies the degree of local pixel intensity variations that occur in the image. A higher contrast value indicates a greater variation in pixel intensity.

- **Correlation** describes the similarity of pixels at specified positions in the image. A high correlation value indicates that the pixels are uniformly distributed in the GLCM, while a low correlation value indicates non-uniform distribution.

- **Dissimilarity** measures the degree of local variation in the image.

- **Energy** provides information about image's homogeneity. Higher energy values indicate a less homogeneous image, while lower energy values indicate a more homogeneous image.
- **Homogeneity** measures the smoothness of the image texture. A higher homogeneity value indicates a smoother texture.

Their mathematical expressions are explicit in Table A.2 of Appendix A.

Table 4.1: Groups of features analysed for each liver segment.

Feature Type	Features
First-order statistical features	Mean
	Variance
	Skewness
	Kurtosis
Second-order statistical features	ASM
	Contrast
	Correlation
	Dissimilarity
	Energy
	Homogeneity

4.2.5 Feature Selection

After extracting all of the features from each liver segment, feature selection was required to determine which features were the most relevant and non-redundant for the classification step. As previously mentioned, the selected features should accurately describe the original dataset while minimising the negative impact of irrelevant features on the model's performance.

So, to reduce the number of features to be used in the classification stage, a ranking-based feature selection method from the Orange platform was used. This is an open source platform commonly used to perform data analysis and visualisation. First, an excel file which contains all the features extracted and corresponding labels (Control and FGR fetus) is loaded on platform. After that, feature selection process is started by using the Rank widget. It scores each feature according to its correlation with the target variable (label) based on a criteria like information gain, ANOVA, chi-square, among others. Next, the best ranked features, i.e the features with the highest score, are selected.

4.2.6 Classification

The current study involved a comparison between different standard supervised ML models, during the performance evaluation. These models were based on the estimators provided by Scikit-learn, a Python package that includes a variety of built-in ML algorithms for the creation

of models. So, after feature extraction and selection stages, five supervised ML models were trained: a DT classifier, a RF classifier, a SVM classifier, a kNN classifier and a NB classifier.

In general, during the process of building any ML model, setting the proper values of the hyperparameters is very important because it may lead to distinct performances. An optimal model architecture can either be achieved by selecting the appropriate algorithm or by obtaining an optimal model architecture. The latter, in turn, is achieved through a hyperparameter tuning (or hyperparameter optimisation) process.

There are many techniques to perform hyperparameter optimisation. However, the grid search approach was the one used for this dissertation. It is a decision-theoretic method that is based on the idea of discovering the best hyperparameter combination within a predefined search space.

The grid search method performs an exhaustive search to find the right combination of hyperparameters that maximises the the performance of the model. At the beginning, a finite set of values for each parameter is given to the grid of configurations. The model is trained upon each possible combination until the best performance results are achieved, which correspond to the optimal configuration. Once the process is finished, the model is fit for training.

Table 4.2: Tested hyperparameter values for each one of the five ML classifiers.

Classifier	Hyperparameters
DT	**criterion** $\in$ {'gini', 'entropy'} **splitter** $\in$ {'best', 'random'} **max_depth** $\in \{3, 6, 9\}$ **min_samples_split** $\in \{10, 20, 30, 40\}$ **min_impurity_decrease** $\in \{0.0, 0.05, 0.1\}$ **random_state** = 42
RF	**n_estimators** $\in \{10, 25, 50, 100, 250, 400\}$ **max_depth** $\in \{3, 6, 9\}$ **criterion** $\in$ {'gini', 'entropy'} **min_samples_split** $\in \{10, 20, 30, 40\}$ **min_samples_leaf** $\in \{5, 10, 15, 20\}$ **random_state** = 42
SVM	**C** $\in \{0.0001, 0.001, 0.01, 0.1, 1, 10, 100\}$ **kernel** $\in$ {'linear', 'poly', 'rbf', 'sigmoid'} **gamma** $\in \{0.0001, 0.001, 0.01, 0.1, 1, 10, 100\}$ **random_state** = 42
kNN	**n_neighbors** $\in \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$ **algorithm** $\in$ {'auto', 'ball_tree', 'kd_tree', 'brute'} **p** $\in \{1, 2\}$
NB	**var_smoothing** $\in \{1.00 \times 10^{-1}, 2.68 \times 10^{-2}, 7.20 \times 10^{-3}, 1.93 \times 10^{-3}, 5.18 \times 10^{-4}, 1.39 \times 10^{-4}, 3.73 \times 10^{-5}, 1.00 \times 10^{-5}, 2.68 \times 10^{-6}, 7.20 \times 10^{-7}, 1.93 \times 10^{-7}, 5.18 \times 10^{-8}, 1.39 \times 10^{-8}, 3.73 \times 10^{-9}, 1.00 \times 10^{-9}\}$

So, as already mentioned, a grid search with leave-one-out cross-validation was performed for models' optimisation due to the small amount of existing data to train the models. Table 4.2 details the values of the hyperparameters tested, and the best hyperparameter configuration for each one of the five ML classifiers is that of the model with the highest accuracy score (view Section 5.3). Hyperparametric tuning proved to be an essential step in ensuring that each model was evaluated under its optimal conditions.

Chapter 5

Results and Discussion

As previously stated, in order to proceed with the study of the echogenicity of the fetal liver, the abdomen of each fetus had to be isolated. In this sense, it was necessary to perform a series of operations, which have already been detailed in the previous chapter, to obtain the intermediate result shown in Figure 5.1. Through the yellow calipers that were already marked on each image it was possible to easily delimit all the images in the HGO database.

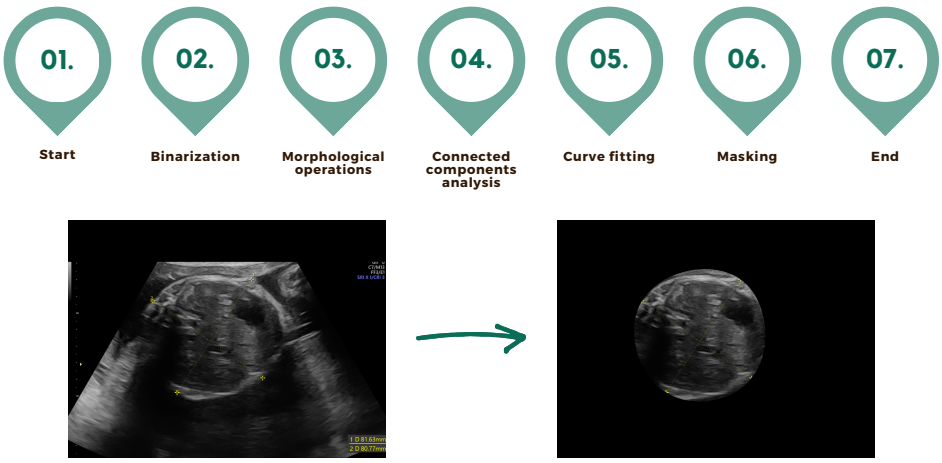

Figure 5.1: Image prepocessing and ROI selection.

The resulting image was used as the basis for extracting features of each liver segment. To obtain these segments, a very intuitive software was designed using PyQt, a library that allows the use of Python's Qt GUI framework. This tool was one of the most important results of the present work whose main window can be seen in Figure 5.2. Despite its functionality, it is still a prototype that should be used as a proof of concept.

5.1 Liver Segmentation - Software

Firstly, the user uploads an US image and then should press the "Segmentation Mode" button to be redirected to a secondary window (Figure 5.3) where it is possible to select a ROI that will later be evaluated. However, this does not always happen. Sometimes, when the user attempts to upload an image and then clicks the "Segmentation Mode" button, a notification appears informing that the file already exists (Figure 5.4). This was done in order to ensure that at the end of the segmentation process there are as many liver segments as input images. If the user wants to start a new study with the same dataset, all segments already created must be deleted. This way, the consistency of the results of feature extraction and classification model selection can be guaranteed.

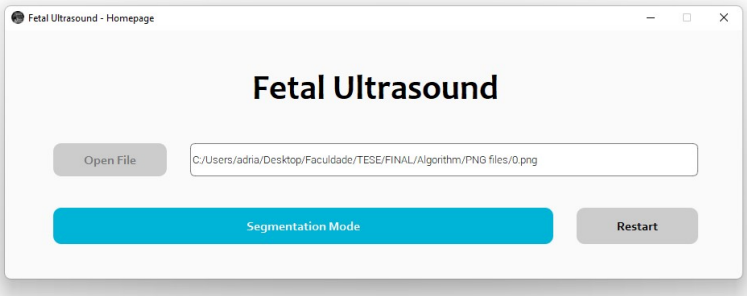

Figure 5.2: Homepage of the developed interface.

Figure 5.3: Secondary window where the user can segment a part of the fetal liver.

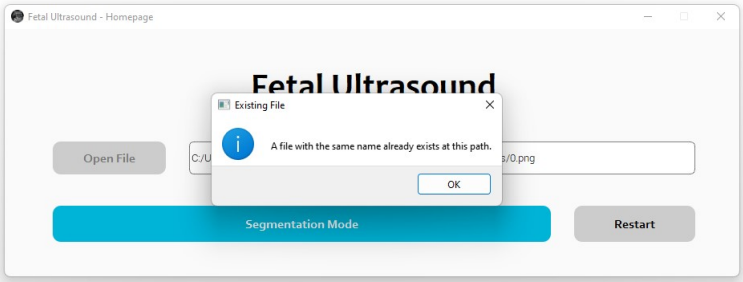

Figure 5.4: Warning message that appears when the user tries to segment an image that has already been segmented and saved.

In the secondary window, the user should set several points along the fetal abdomen. The inserted points, marked in yellow as shown in Figure 5.5, are connected by a blue line that shapes the outline of the liver segment. When the "Accept" button is pressed, the selection is saved, and the user backs to the main window. During this procedure, the delineation might be performed multiple times until the user feels confident that the contour was properly executed. For this purpose, a "Clear Path" button was included. When the user needs to restart the delineation process, this button should be pressed and another information message should appear on the screen (Figure 5.6). Additionally, a "Zoom" button has been added to this window in an effort to improve the precision of the delineation process.

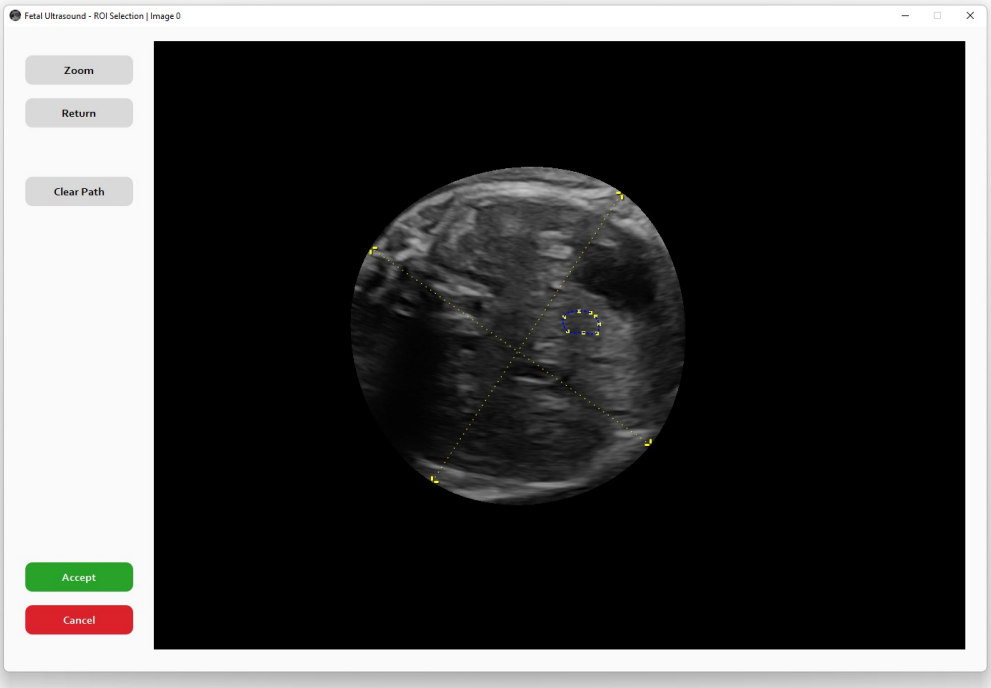

Figure 5.5: Example of the liver delineation that will be considered for textural analysis.

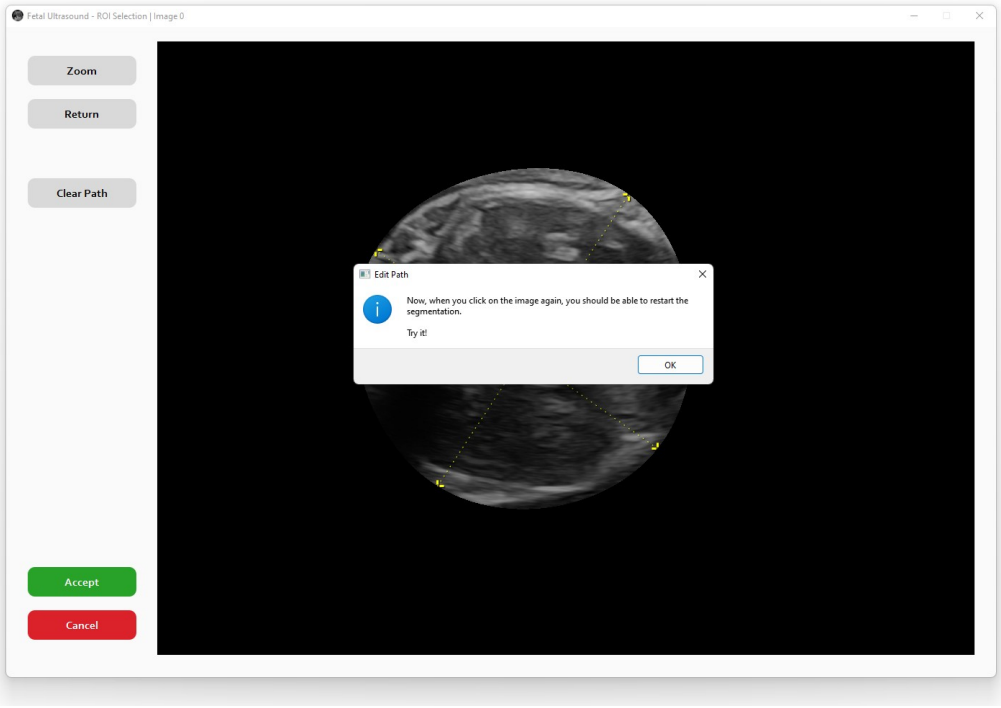

Figure 5.6: Message that appears when the user needs to restart the delineation process.

5.2 Feature Extraction and Feature Selection

A feature analysis was performed by calculating histogram features and GLCM features for each liver segment. A total of twenty-eight features per liver segment was extracted and grouped as follows: four histogram features and twenty-four GLCM features since all available dimensions were considered. All this information is organised in Table 5.1.

After that feature selection was conducted. According what was previously mentioned in Chapter 4.2.5, it was performed using the Orange tool to reduce the number of features to be used in the classification stage. To ascertain the relevance of each feature, the scoring criteria chosen was the ANOVA criteria since it is widely used when the input variables are numeric and the target variable is categorical. In regards to the amount of features selected, the nine features with the highest score were considered, accounting for one-third of the total number of features extracted. The features selected by the Rank widget were ASM_45, ASM_90, correlation_0, correlation_90, correlation_135, energy_45, homogeneity_45, homogeneity_90 and skewness. These findings suggest that attributes such as the distribution of grey levels, the degree of pixel similarity at specific positions, and the smoothness of image texture may play an important role in distinguishing the fetus. As mentioned in section 2.3.5, the fetal liver typically appears hypoechoic in relation to the ribs and lungs, and hyperechoic in relation to the umbilical vein. However, with the existing data, it remains unclear whether these patterns of echogenicity can distinguish fetuses with and without FGR.

In order to verify these results, a transform method from the Scikit-learn Python library was also used to select the best features based on univariate statistical tests. The method used

was the *SelectKBest* and the *f_classif* function, which calculates the ANOVA F-value for the provided sample, was passed in the *score_func* parameter. The results obtained are illustrated in Figure 5.7 and are in agreement with the previous ones.

Table 5.1: Extracted features for each fetal liver segment.

Feature Type	**Number of Features**	**Features**	
First-order statistical features	4	mean	
		variance	
		skewness	
		kurtosis	
Second-order statistical features	24	Angular Second Moment	ASM_0
			ASM_45
			ASM_90
			ASM_135
		Contrast	contrast_0
			contrast_45
			contrast_90
			contrast_135
		Correlation	correlation_0
			correlation_45
			correlation_90
			correlation_135
		Dissimilarity	dissimilarity_0
			dissimilarity_45
			dissimilarity_90
			dissimilarity_135
		Energy	energy_0
			energy_45
			energy_90
			energy_135
		Homogeneity	homogeneity_0
			homogeneity_45
			homogeneity_90
			homogeneity_135

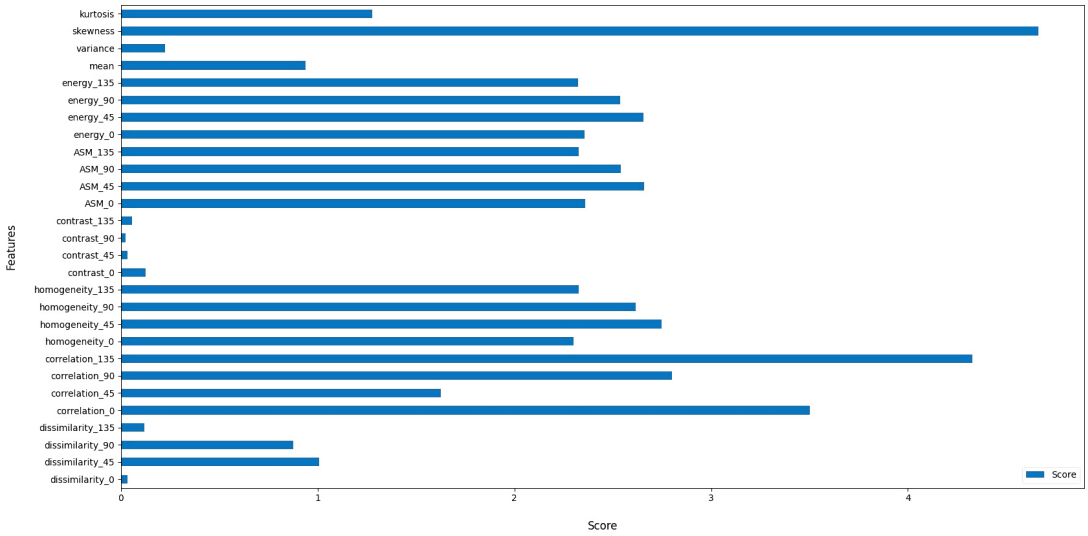

Figure 5.7: Extracted features importance.

Following that, an Exploratory Data Analysis (EDA) was performed with the purpose of doing a more visual analysis of the entire dataset before making any assumptions. This strategy not only helps to identify obvious errors, but it also helps to better understand certain patterns within the data, detect outliers and find interesting relationships between variables.

The EDA was accomplished with Sweetviz, one of the most recent open-source Python libraries, where a self-contained HyperText Markup Language (HTML) report may be generated from a pandas dataframe. As such, first the HGO dataset was divided into two different subgroups, one for fetuses without growth restriction (hereinafter referred to as the control group) and another for fetuses with FGR (hereinafter referred to as the FGR group), which were supplied as inputs to the *sweetviz.compare()* function. Then, the HTML report was generated, incorporating univariate analysis of each feature, to see whether the nine aforementioned features are sufficiently discriminating, i.e. whether they successfully differentiate one group from another.

The results are presented in Figure 5.8 which contains nine graphs showing the distribution of FGR group (blue bars) and control group (orange bars). Overall, there is no missing data and some outliers were identified, specially in the distribution of non-FGR fetuses. The distribution of FGR fetuses is typically shifted to the left relative to the distribution of fetuses from control group, except in *correlation_0*, *correlation_90* and *correlation_135* cases. However, the distribution of both groups is nearly overlapping, resulting in the finding that no one single feature can accurately differentiate the fetuses.

Nonetheless, it was decided to build the ML model based on the aforementioned nine features, in the expectation that all of them would be able to accomplish the classification task efficiently.

Figure 5.8: EDA results only for the nine selected features. The blue bars represent the distribution of the FGR group, whereas the orange bars represent the distribution of the control group.

5.3 Model Evaluation

The process of selecting a final ML model for a given data set is known as model selection. It can be applied to models of the same type configured with different hyperparameters, as well as models of different types. Thus, the five ML classifiers were tested for a total of 43 US images from the HGO database. The optimal architecture of each model was achieved automatically after hyperparameter tuning, whose results are specified in Table 5.2. For that purpose, the *model_selection.GridSearchCV* module of the scikit-learn Python library was used, as well as the *tree.DecisionTreeClassifier*, *ensemble.RandomForestClassifier*, *svm.SVC*, *neighbors.KNeighborsClassifier* and *naive_bayes.GaussianNB* modules. The former performs hyperparameter tuning through a grid search strategy, whereas the others are the classifiers themselves.

Table 5.2: Optimal hyperparameter values for each classifier.

Classifier	Hyperparameters
DT	**criterion** = 'entropy' **splitter** = 'random' **max_depth** = 3 **min_samples_split** = 10 **min_impurity_decrease** = 0.0 **random_state** = 42
RF	**n_estimators** = 10 **criterion** = 'gini' **max_depth** = 3 **min_samples_split** = 10 **min_samples_leaf** = 10 **random_state** = 42
SVM	**C** = 1.0 **kernel** = 'rbf' **gamma** = 100 **random_state** = 42
kNN	**n_neighbors** = 9 **algorithm** = 'auto' **p** = 1
NB	**var_smoothing** = 0.1

This first selection allowed obtaining the optimal set of hyperparameters for the five different types of ML models. However, it is still required to evaluate which of these models is more suitable for the given problem.

In an ideal situation, where there is a significant amount of data, the initial dataset should be divided into training, validation and test sets. Then, it is essential to fit the available ML models on the training set, evaluate and select the best-performing one on the validation set, and then report its performance on the test set. In this study, as the available dataset is quite limited, this approach had to be replaced by a resampling model selection method, such as cross-validation.

Resampling methods split the training dataset into subsets for training and testing. The ML model should be fitted and evaluated on the subset available for training and testing, respectively. This process may be repeated multiple times and the model's performance on each iteration should be reported. The final model performance is given by the average of each individual score. This kind of approach improves the stability of ML algorithms, as well as helps in avoiding overfitting.

The preferred cross-validation method for this study was the leave-one-out where only one sample is used for testing, while the rest are used to train the model. So, given that the dataset has 43 samples, it means that 43 train and test splits of the dataset were created, with each single row of the dataset given an opportunity to be used as the test set. Similarly, 43 models were created and evaluated.

The performance of the classifiers was assessed based on the metrics mentioned in Chapter 2.4.4. In this binary classification problem, the positive class is represented by the fetuses with FGR (hence, FGR group) and the negative class is represented by the fetuses with normal fetal growth (hence, control group). Thus, the following relationships were established:

- **TP:** number of fetuses with FGR that the model correctly classifies.
- **FP:** number of fetuses with normal fetal growth that the model classifies as fetuses with FGR. These cases should be interpreted as fetuses with possible FGR and as such need to be examined by specialists.
- **TN:** number of fetuses with normal fetal growth that the model correctly classifies.
- **FN:** number of fetuses with FGR that the model classifies as fetuses with normal fetal growth. Clinically, FNs are a cause for concern because it means that these fetuses will not receive appropriate treatment, which may aggravate their clinical condition.

So, in a medical context, FPs may increase the cost of FGR detection, whereas FNs may delay treatment to the point where it is no longer effective.

Figure 5.9 displays the confusion matrices obtained for the five aforementioned classifiers. Table 5.3 exhibits all the metrics that can be computed from these matrices and will be used to assess the performance of each model.

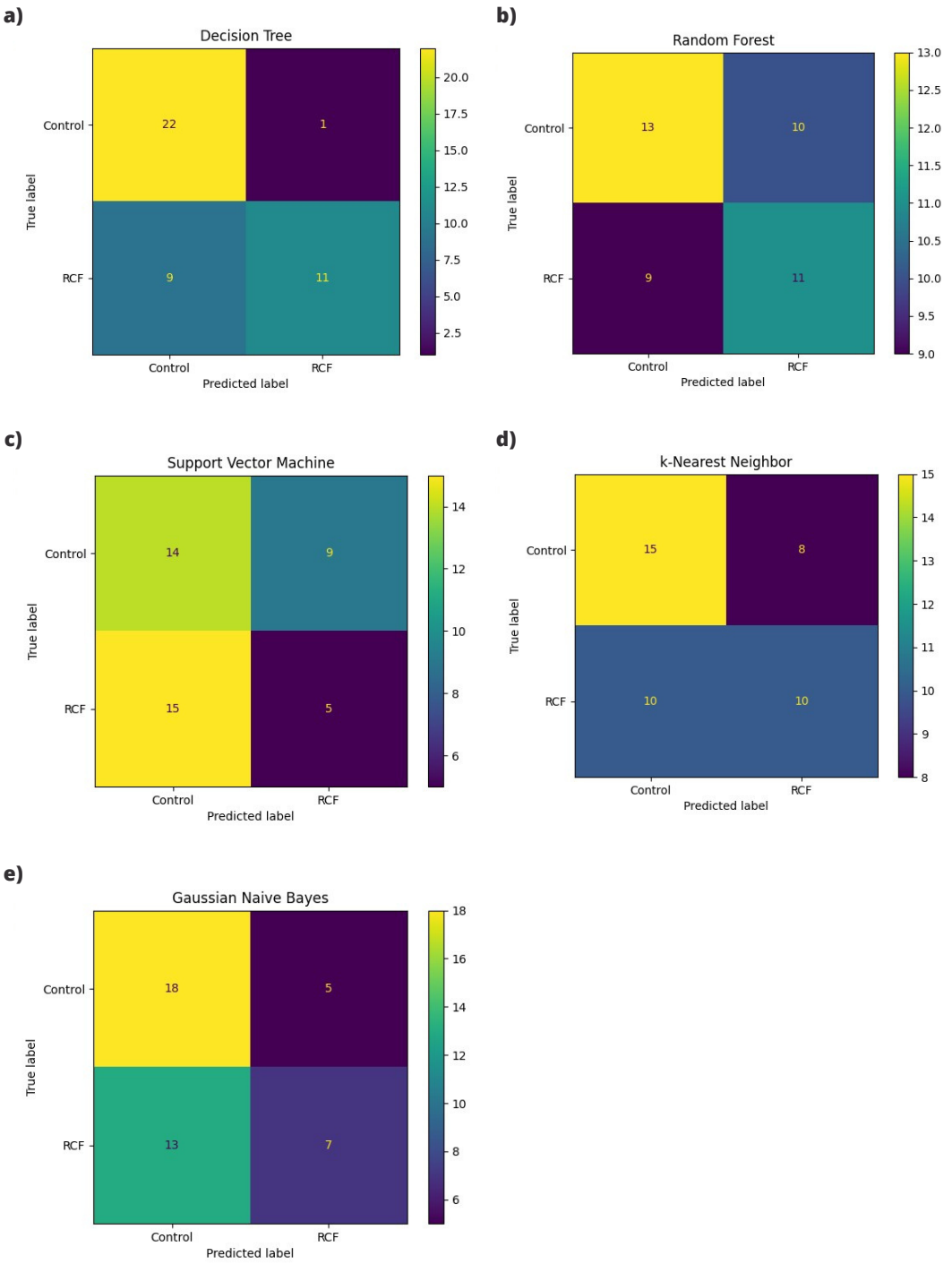

Figure 5.9: Confusion matrices of the classification stage obtained for a) DT classifier, b) RF classifier, c) SVM classifier, d) kNN classifier, e) NB classifier.

Table 5.3: Performance evaluation metrics achieved by the five ML models.

Classifier	Accuracy	Precision	Recall	F1-Score	Specificity	AUROC
DT	0.7674	0.9167	0.5500	0.6875	0.9565	0.6152
RF	0.5581	0.5238	0.5500	0.5366	0.5652	0.4652
SVM	0.4419	0.3571	0.2500	0.2941	0.6087	0.6152
kNN	0.5814	0.5556	0.5000	0.5263	0.6522	0.5054
NB	0.5814	0.5833	0.3500	0.4375	0.7826	0.5065

A high-precision model is a model that correctly classifies fetuses with FGR only a few times, but when it does it is possible to ensure that it is indeed a FGR fetus. Thus, the model is able to correctly distinguish a fetus with FGR from a fetus with normal fetal growth when it is in fact a fetus with FGR. On the other hand, a model with a greater recall means it will be able to recognise as many FGR fetuses as possible. Thus, it is possible to be sure that all fetuses with an abnormal fetal growth that are identified by the model are subjected to medical examination. When recall is used to assess the performance of the model, it is often compared to specificity. This is a performance metric that measures the proportion of fetuses without FGR that are correctly identified by the model. So, a model with a high specificity is a model that correctly predicts the majority of fetuses without FGR.

Precision and recall are two metrics that are mutually exclusive. Therefore, it is impossible to have a model with high values of both metrics. According to what has previously been stated, the final model should aim to reduce the risk of FNs. This means that the recall measure should be as near to 100% as possible, since this means that all cases of fetuses with FGR have been correctly predicted (i.e., $FN = 0$). So, in a medical context, recall metric is more important than precision.

Although accuracy is one key measure for classification problems, it is not always the most reliable. This metric deals only with TP and TN cases, but as noted above, the distribution of FP and specially FN is very important for this study. On the other hand, the AUROC metric is widely used in medical research since high AUROC values mean that the model under evaluation classifies more instances correctly. Moreover, this metric is one of the most widely used when evaluating different learning algorithms because it takes into account all class distributions, as well as the error costs of the model. In the current study, the higher the AUROC, the better the model's ability to distinguish between fetuses with FGR and fetuses with normal fetal growth.

According to the values in Table 5.3 is possible to conclude that the DT model is the most accurate, with an accuracy of 0.7674 compared to the values achieved by the other models ranging from 0.4419 to 0.5814. This model was also the one that presented the highest precision, with a value of 0.9167 which differs greatly from the values obtained by the remaining models. The recall metric was the most consistent, since the values are relatively close to each other. Still, the DT and RF models were the ones that stood out, both with a value of 0.5500.

On the other hand, it is possible to identify two models that present a greater specificity: DT and NB with a value of 0.9565 and 0.7826, respectively. Finally, the AUROC values show that the kNN and NB models cannot differentiate between FGR and non-FGR fetuses since their AUROC is close to 0.5. This means that these models either predict a random class or a constant class for all instances. So, they are very unreliable. Conversely, the DT and SVM models, both with an AUROC of 0.6152, are more likely to distinguish between the two classes because they are able to detect more TP and TN cases than FP and FN cases.

The ROC curves for each of the five models of ML are illustrated in Figure 5.10, Figure 5.11, Figure 5.12, Figure 5.13 and Figure 5.14, and provide evidence of the relationships established above.

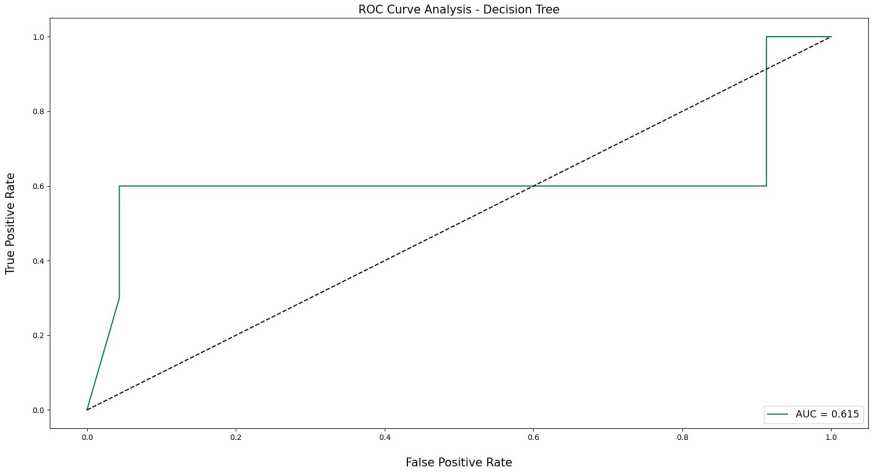

Figure 5.10: ROC curve for DT classifier.

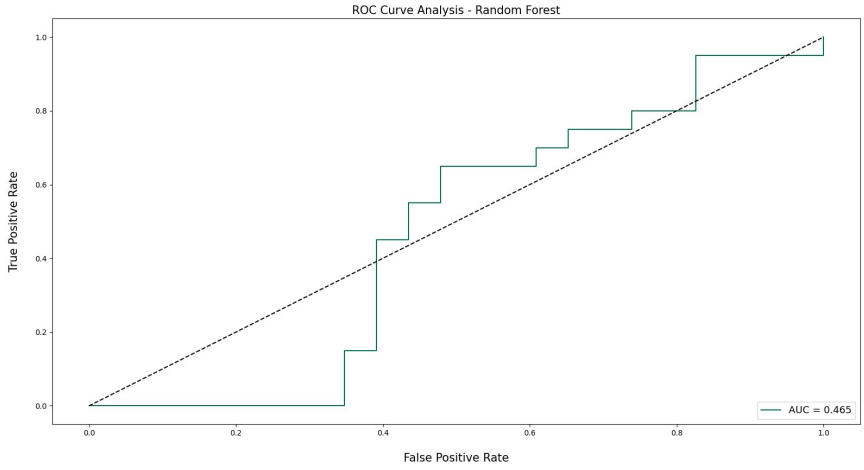

Figure 5.11: ROC curve for RF classifier.

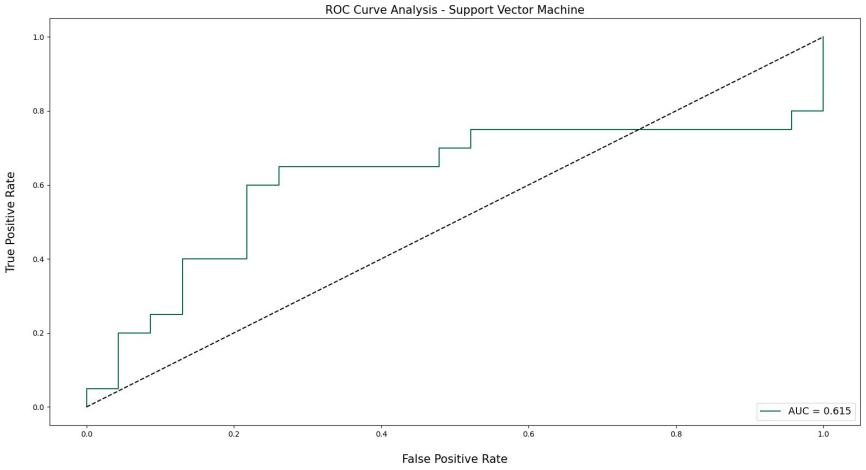

Figure 5.12: ROC curve for SVM classifier.

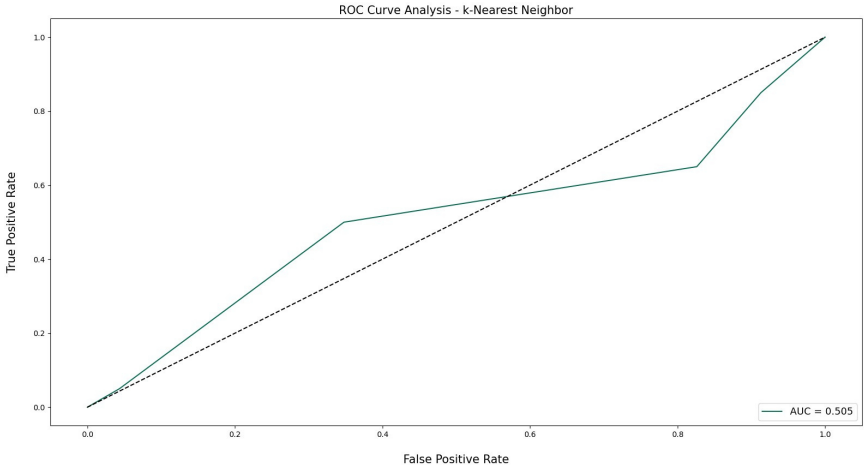

Figure 5.13: ROC curve for kNN classifier.

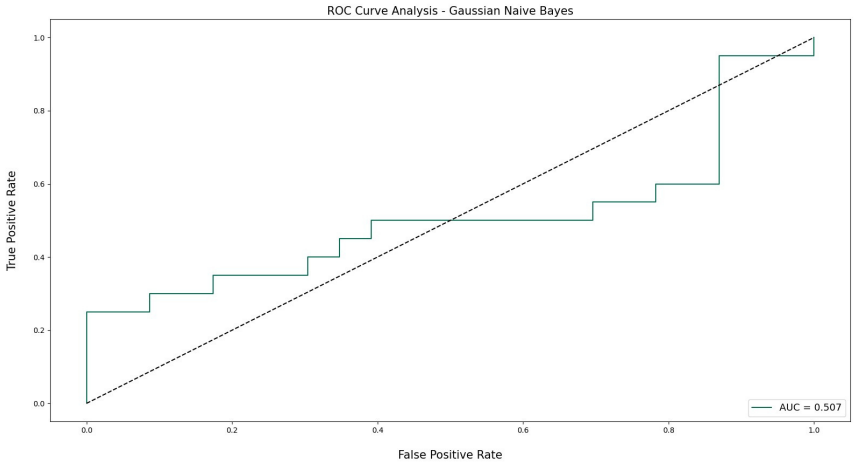

Figure 5.14: ROC curve for NB classifier.

By observing the classification results for the five different ML models, it is possible to conclude that DT performs remarkably better. Overall, this model achieved the best results for all performance metrics. This has particular relevance in the medical context, where metrics such as recall and AUROC must have higher values. However, as real-time classification was not performed due to lack of data, the results achieved by the different models may be a consequence of overfitting.

CHAPTER 6

CONCLUSION

Fetal Growth Restriction (FGR) is a common complication of pregnancy that has been associated with some negative perinatal outcomes, with fetuses being predisposed to the onset of a wide variety of diseases. It is the second leading cause of perinatal mortality and one of the most common causes of preterm births. All this leads to the need for early diagnosis of this medical condition.

Although the application of AI in the fields of obstetrics and gynaecology is still in the early stages, several studies have been proposed for automating the segmentation and measurement of fetal biometry. Furthermore, some studies have also been published for the detection of growth patterns and anomalies, such as FGR. However, an extensive literature search revealed that to date there have been no studies that have attempted to correlate potential changes in fetal liver echogenicity with the FGR.

Therefore, this dissertation proposed a novel approach to assess the feasibility of relating fetal liver echogenicity to the diagnosis of FGR using image processing techniques and AI in 43 US images from the HGO database. To this purpose, a method was developed that attempted to isolate the fetal abdomen in order to aid the liver's segmentation. Overall, for the delimitation of fetal abdomen it was necessary to perform a connected component analysis, apply several morphological operations, as well as delineate the ROI using a curve fitting approach. Then, the resulting image was used in the developed tool to obtain a segment of the fetal liver. This tool was designed to enable the manual selection of the segment, since it was believed that it would be possible to perform a reliable analysis of fetal liver echogenicity through a small portion of this organ. This tool presents only the necessary functionalities to perform the intended task, so it should be seen as a prototype and, as such, used as a proof of concept.

Next, a textural analysis of these segments was performed, which allowed extracting a total of twenty-eight features of which nine were effectively used in the classification phase. However, prior to that an EDA was performed which demonstrated that these features do not

appear to be the most adequate for the diagnosis of FGR, since they appear to be unable to distinguish fetuses with FGR from non-restricted fetuses. Nonetheless, it was decided to keep the initial approach in the expectation that all nine together would be able to accomplish the classification task efficiently. In order to conduct a classification evaluation, five ML models were trained and their performance was assessed through different evaluation metrics to conclude which model was the most appropriate to diagnose FGR from fetal liver analysis. From among the five models tested, the one that produced the best results was DT (accuracy = 0.7674, precision = 0.9167, recall = 0.5500, specificity = 0.9565 and AUROC = 0.6152).

The results obtained in this dissertation are not very conclusive because the dataset used is quite limited and ML algorithms require a large amount of data to learn and work properly, regardless of the complexity of the problem. Besides that, the dataset is also unrepresentative. In a real medical context, the number of fetuses with normal fetal growth is much higher than the number of fetuses with FGR, and this is not the case in this dissertation.

Thus, responding to the hypothesis initially stated: it is currently impossible to ensure that changes in fetal liver echogenicity are directly related to the diagnosis of FGR. Nevertheless, the results obtained do not deny the feasibility of this correlation which will necessarily require further studies to obtain definitive conclusions.

6.1 Future Work

Some potential solutions to enhance the work done in this dissertation are presented below:

- Increase the database in a larger range of images: This work should have been done using a more extensive database of ultrasound images. As such, the real-time classification component was missing since the few existing data were used to train the model.
- Automatic segmentation of the fetal abdomen (new attempt): Some research have demonstrated that it is a feasible approach, so it should be considered in a future work.
- Implementation of more features in the developed tool: Create a section dedicated to image preprocessing, improve the "Zoom" option to enhance segmentation accuracy and include a component to classify new exams are some examples of potential new features.
- Consider extracting new features: The features extracted in this dissertation seem to be unsuitable for FGR diagnosis.

Bibliography

[1] Z. Chen, Z. Liu, M. Du, and Z. Wang, "Artificial Intelligence in Obstetric Ultrasound: An Update and Future Applications", *Frontiers in Medicine*, vol. 8, 2021. DOI: 10.3389/fmed.2021.733468 (cit. on pp. 2, 16).

[2] W. W. Hay, "Growth and Development: Physiological Aspects", in *Encyclopedia of Human Nutrition*, B. Caballero, Ed., 3rd ed. Waltham, MA, USA: Academic Press, 2013, pp. 399–407, ISBN: 978-0-12-384885-7. DOI: 10.1016/B978-0-12-375083-9.00139-2 (cit. on pp. 4–6).

[3] F. Gaccioli and S. Lager, "Placental Nutrient Transport and Intrauterine Growth Restriction", *Frontiers in Physiology*, vol. 7, no. 40, 2016. DOI: 10.3389/fphys.2016.00040 (cit. on p. 4).

[4] L. M. M. Nardozza, A. C. R. Caetano, A. C. P. Zamarian, *et al.*, "Fetal growth restriction: Current knowledge", *Archives of Gynecology and Obstetrics*, vol. 295, no. 5, pp. 1061–1077, 2017. DOI: 10.1007/s00404-017-4341-9 (cit. on pp. 4, 5, 7, 9).

[5] P. N. S. Rao, A. Shashidhar, and C. Ashok, "In utero fuel homeostasis: Lessons for a clinician", *Indian Journal of Endocrinology and Metabolism*, vol. 17, no. 1, pp. 60–68, 2013. DOI: 10.4103/2230-8210.107851 (cit. on pp. 5, 6).

[6] N. Melamed, A. Baschat, Y. Yinon, *et al.*, "FIGO (International Federation of Gynecology and Obstetrics) initiative on fetal growth: Best practice advice for screening, diagnosis, and management of fetal growth restriction", *International Journal of Gynecology & Obstetrics*, vol. 152, no. S1, pp. 3–57, 2021. DOI: 10.1002/ijgo.13522 (cit. on pp. 6–9, 23).

[7] S. J. Gordijn, I. M. Beune, B. Thilaganathan, *et al.*, "Consensus definition of fetal growth restriction: a Delphi procedure", *Ultrasound in Obstetrics & Gynecology*, vol. 48, no. 3, pp. 333–339, 2016. DOI: 10.1002/uog.15884 (cit. on pp. 7, 8).

[8] B. A. Campbell, "Utilizing sonography to follow fetal growth", *Obstetrics and Gynecology Clinics of North America*, vol. 25, no. 3, pp. 597–607, 1998. DOI: 10.1016/S0889-8545(05)70029-2 (cit. on p. 8).

[9] D. Sharma, N. Farahbakhsh, S. Shastri, and P. Sharma, "Intrauterine growth restriction – part 2", *The Journal of Maternal-Fetal & Neonatal Medicine*, vol. 29, no. 24, pp. 4037–4048, 2016. DOI: 10.3109/14767058.2016.1154525 (cit. on pp. 8, 10, 11).

[10] E. Platz and R. Newman, "Diagnosis of IUGR: Traditional Biometry", *Seminars in Perinatology*, vol. 32, no. 3, pp. 140–147, 2008. DOI: 10.1053/j.semperi.2008.02.002 (cit. on p. 8).

[11] J. S. Dashe, D. D. McIntire, M. J. Lucas, and K. J. Leveno, "Effects of symmetric and asymmetric fetal growth on pregnancy outcomes", *Obstetrics & Gynecology*, vol. 96, no. 3, pp. 321–327, 2000. DOI: 10.1016/S0029-7844(00)00943-1 (cit. on p. 8).

[12] F. Figueras and E. Gratacos, "Stage-based approach to the management of fetal growth restriction", *Prenatal Diagnosis*, vol. 34, no. 7, pp. 655–659, 2014. DOI: 10.1002/pd.4412 (cit. on p. 8).

[13] S. Savchev, F. Figueras, M. Sanz-Cortes, *et al.*, "Evaluation of an Optimal Gestational Age Cut-Off for the Definition of Early- and Late-Onset Fetal Growth Restriction", *Fetal Diagnosis and Therapy*, vol. 36, no. 2, pp. 99–105, 2014. DOI: 10.1159/000355525 (cit. on p. 8).

[14] A. Dall'Asta, V. Brunelli, F. Prefumo, T. Frusca, and C. C. Lees, "Early onset fetal growth restriction", *Maternal Health, Neonatology and Perinatology*, vol. 3, pp. 1–12, 2017. DOI: 10.1186/s40748-016-0041-x (cit. on p. 8).

[15] F. Figueras, J. Caradeux, F. Crispi, E. Eixarch, A. Peguero, and E. Gratacos, "Diagnosis and surveillance of late-onset fetal growth restriction", *American Journal of Obstetrics & Gynecology*, vol. 218, no. 2, S790–S802, 2018. DOI: 10.1016/j.ajog.2017.12.003 (cit. on p. 8).

[16] F. Figueras, E. Eixarch, E. Meler, *et al.*, "Small-for-gestational-age fetuses with normal umbilical artery Doppler have suboptimal perinatal and neurodevelopmental outcome", *European Journal of Obstetrics and Gynecology and Reproductive Biology*, vol. 136, no. 1, pp. 34–38, 2008. DOI: 10.1016/j.ejogrb.2007.02.016 (cit. on p. 9).

[17] F. M. Severi, C. Bocchi, A. Visentin, *et al.*, "Uterine and fetal cerebral Doppler predict the outcome of third-trimester small-for-gestational age fetuses with normal umbilical artery Doppler", *Ultrasound in Obstetrics & Gynecology*, vol. 19, no. 3, pp. 225–228, 2002. DOI: 10.1046/j.1469-0705.2002.00652.x (cit. on p. 9).

[18] D. Sharma, S. Shastri, N. Farahbakhsh, and P. Sharma, "Intrauterine growth restriction – part 1", *The Journal of Maternal-Fetal & Neonatal Medicine*, vol. 29, no. 24, pp. 3977–3987, 2016. DOI: 10.3109/14767058.2016.1152249 (cit. on p. 9).

[19] A. A. Mahendru, A. Daemen, T. R. Everett, *et al.*, "Impact of ovulation and implantation timing on first-trimester crown–rump length and gestational age", *Ultrasound in Obstetrics & Gynecology*, vol. 40, no. 6, pp. 630–635, 2012. DOI: 10.1002/uog.12277 (cit. on p. 9).

[20] F. P. Hadlock, Y. P. Shah, D. J. Kanon, and J. V. Lindsey, "Fetal crown-rump length: reevaluation of relation to menstrual age (5-18 weeks) with high-resolution real-time US", *Radiology*, vol. 182, no. 2, pp. 501–505, 1992. DOI: 10.1148/radiology.182.2.1732970 (cit. on p. 9).

[21] S. Daya, "Accuracy of gestational age estimation by means of fetal crown-rump length measurement", *American Journal of Obstetrics & Gynecology*, vol. 168, no. 3, pp. 903–908, 1993. DOI: 10.1016/s0002-9378(12)90842-x (cit. on p. 9).

[22] D. A. Savitz, J. Terry James W., N. Dole, J. Thorp John M., A. M. Siega-Riz, and A. H. Herring, "Comparison of pregnancy dating by last menstrual period, ultrasound scanning, and their combination", *American Journal of Obstetrics & Gynecology*, vol. 187, no. 6, pp. 1660–1666, 2002. DOI: 10.1067/mob.2002.127601 (cit. on p. 9).

[23] N. Melamed, Y. Yogev, I. Meizner, R. Mashiach, R. Bardin, and A. Ben-Haroush, "Sonographic Fetal Weight Estimation", *Journal of Ultrasound in Medicine*, vol. 28, no. 5, pp. 617–629, 2009. DOI: 10.7863/jum.2009.28.5.617 (cit. on p. 9).

[24] N. J. Dudley, "A systematic review of the ultrasound estimation of fetal weight", *Ultrasound in Obstetrics & Gynecology*, vol. 25, no. 1, pp. 80–89, 2005. DOI: https://doi.org/10.1002/uog.1751 (cit. on p. 9).

[25] J. Milner and J. Arezina, "The accuracy of ultrasound estimation of fetal weight in comparison to birth weight: A systematic review", *Ultrasound*, vol. 26, no. 1, pp. 32–41, 2018. DOI: 10.1177/1742271X17732807 (cit. on pp. 9, 24).

[26] A. C. of Obstetricians, Gynecologists, *et al.*, "ACOG Practice Bulletin No. 204: Fetal Growth Restriction", *Obstetrics & Gynecology*, vol. 133, no. 2, e97–e109, 2019. DOI: 10.1097/AOG.0000000000003070 (cit. on p. 10).

[27] J. Gardosi, V. Madurasinghe, M. Williams, A. Malik, and A. Francis, "Maternal and fetal risk factors for stillbirth: Population based study", *BMJ*, vol. 346, 2013. DOI: 10.1136/bmj.f108 (cit. on p. 10).

[28] V. Flenady, L. Koopmans, P. Middleton, *et al.*, "Major risk factors for stillbirth in high-income countries: A systematic review and meta-analysis", *The Lancet*, vol. 377, no. 9774, pp. 1331–1340, 2011. DOI: 10.1016/S0140-6736(10)62233-7 (cit. on p. 10).

[29] A. Rosenberg, "The IUGR Newborn", *Seminars in Perinatology*, vol. 32, no. 3, pp. 219–224, 2008. DOI: https://doi.org/10.1053/j.semperi.2007.11.003 (cit. on p. 10).

[30] E. K. Pallotto and H. W. Kilbride, "Perinatal Outcome and Later Implications of Intrauterine Growth Restriction", *Clinical Obstetrics and Gynecology*, vol. 49, no. 2, pp. 257–269, 2006. DOI: 10.1097/00003081-200606000-00008 (cit. on p. 10).

[31] D. J. P. Barker, "Adult Consequences of Fetal Growth Restriction", *Clinical Obstetrics and Gynecology*, vol. 49, no. 2, pp. 270–283, 2006. DOI: 10.1097/00003081-200606000-00009 (cit. on p. 10).

[32] I. Edler and K. Lindström, "The history of echocardiography", *Ultrasound in Medicine & Biology*, vol. 30, no. 12, pp. 1565–1644, 2004. DOI: 10.1016/S0301-5629(99)00056-3 (cit. on p. 11).

[33] R. Abinader and S. L. Warsof, "Benefits and Pitfalls of Ultrasound in Obstetrics and Gynecology", *Obstetrics and Gynecology Clinics of North America*, vol. 46, no. 2, pp. 367–378, 2019. DOI: 10.1016/j.ogc.2019.01.011 (cit. on pp. 11, 16).

[34] V. Chan and A. Perlas, "Basics of Ultrasound Imaging", in *Atlas of Ultrasound-Guided Procedures in Interventional Pain Management*, S. N. Narouze, Ed. New York, NY: Springer New York, 2011, ch. 2, pp. 13–19, ISBN: 978-1-4419-1681-5. DOI: 10.1007/978-1-4419-1681-5_2 (cit. on pp. 11–14).

[35] A. Carovac, F. Smajlovic, and D. Junuzovic, "Application of Ultrasound in Medicine", *Acta Informatica Medica*, vol. 19, no. 3, pp. 168–171, 2011. DOI: 10.5455/aim.2011.19.168-171 (cit. on p. 11).

[36] S. Abbas and P. Peng, "Basic Principles and Physics of Ultrasound", in *Ultrasound for Interventional Pain Management: An Illustrated Procedural Guide*, P. Peng, R. Finlayson, S. H. Lee, and A. Bhatia, Eds. Cham: Springer International Publishing, 2020, ch. 1, pp. 1–31, ISBN: 978-3-030-18371-4. DOI: 10.1007/978-3-030-18371-4_1 (cit. on pp. 11–15).

[37] A. Abuhamad, E. Walsh, R. Chaoui, P. Jeanty, and D. Paladini, "Basic Physical Principles of Medical Ultrasound", in *Ultrasound in Obstetrics and Gynecology: A Practical Approach*, 1st ed. 2014, ch. 1, pp. 9–29 (cit. on pp. 11, 12, 15).

[38] H. Sehmbi and A. Perlas, "Basics of Ultrasound Imaging", in *Regional Nerve Blocks in Anesthesia and Pain Therapy: Traditional and Ultrasound-Guided Techniques*, P. Peng, R. Finlayson, S. H. Lee, and A. Bhatia, Eds. Cham: Springer International Publishing, 2015, ch. 3, pp. 27–56, ISBN: 978-3-319-05131-4. DOI: 10.1007/978-3-319-05131-4_3 (cit. on pp. 11–15).

[39] J. E. Aldrich, "Basic physics of ultrasound imaging", *Critical care medicine*, vol. 35, no. 5, S131–S137, 2007. DOI: 10.1097/01.CCM.0000260624.99430.22 (cit. on p. 11).

[40] R. M. Rangayyan, "The Nature of Biomedical Images", in *Biomedical Image Analysis*, 1st ed. Boca Raton, FL, US: CRC Press, 2004, ch. 1, pp. 1–59 (cit. on p. 11).

[41] B. Ihnatsenka and A. P. Boezaart, "Ultrasound: Basic understanding and learning the language", *International Journal of Shoulder Surgery*, vol. 4, no. 3, pp. 55–62, 2010. DOI: 10.4103/0973-6042.76960 (cit. on p. 15).

[42] Z. Papp and T. Fekete, "The evolving role of ultrasound in obstetrics/gynecology practice", *International Journal of Gynecology & Obstetrics*, vol. 82, no. 3, pp. 339–346, 2003. DOI: 10.1016/S0020-7292(03)00224-8 (cit. on p. 16).

[43] Committee on Practice Bulletins - Obstetrics and the American Institute of Ultrasound in Medicine, "Practice Bulletin No. 175: Ultrasound in Pregnancy", *Obstetrics & Gynecology*, vol. 128, no. 6, e241–e256, 2016. DOI: 10.1097/AOG.0000000000001815 (cit. on p. 16).

[44] L. J. Salomon, Z. Alfirevic, C. Bilardo, *et al.*, "ISUOG Practice Guidelines: performance of first-trimester fetal ultrasound scan", *Ultrasound in Obstetrics & Gynecology*, vol. 41, no. 1, pp. 102–113, 2013. DOI: 10.1002/uog.12342 (cit. on p. 16).

[45] L. J. Salomon, Z. Alfirevic, V. Berghella, *et al.*, "Practice guidelines for performance of the routine mid-trimester fetal ultrasound scan", *Ultrasound in Obstetrics & Gynecology*, vol. 37, no. 1, pp. 116–126, 2011. DOI: 10.1002/uog.8831 (cit. on pp. 16, 30).

[46] D. Arduini and F. Giacomello, "Fetal biometry, estimation of gestational age, assessment of fetal growth", in *Ultrasound in Obstetrics and Gynaecology*, J. W. Wladimiroff and S. H. Eik-Nes, Eds. Edinburgh: Elsevier, 2009, ch. 9, pp. 141–155, ISBN: 978-0-444-51829-3. DOI: 10.1016/B978-0-444-51829-3.00009-X (cit. on p. 17).

[47] B. Žaliūnas, D. Bartkevičienė, G. Drąsutienė, A. Utkus, and J. Kurmanavičius, "Fetal biometry: Relevance in obstetrical practice", *Medicina*, vol. 53, no. 6, pp. 357–364, 2017. DOI: 10.1016/j.medici.2018.01.004 (cit. on pp. 17, 18).

[48] R. J. M. Snijders and K. H. Nicolaides, "Fetal biometry at 14–40 weeks' gestation", *Ultrasound in Obstetrics & Gynecology*, vol. 4, no. 1, pp. 34–48, 1994. DOI: 10.1046/j.1469-0705.1994.04010034.x (cit. on p. 17).

[49] T. C. Chang, S. C. Robson, R. J. Boys, and J. A. Spencer, "Prediction of the small for gestational age infant: Which ultrasonic measurement is best?", *Obstetrics and Gynecology*, vol. 80, no. 6, pp. 1030–1038, 1992. [Online]. Available: http://europepmc.org/abstract/MED/1448248 (cit. on p. 17).

[50] C. Janiesch, P. Zschech, and K. Heinrich, "Machine learning and deep learning", *Electronic Markets*, vol. 31, no. 3, pp. 685–695, 2021. DOI: 10.1007/s12525-021-00475-2 (cit. on p. 18).

[51] M. A. El Mrabet, K. El Makkaoui, and A. Faize, "Supervised Machine Learning: A Survey", in *2021 4th International Conference on Advanced Communication Technologies and Networking (CommNet)*, IEEE, 2021, pp. 1–10. DOI: 10.1109/CommNet52204.2021.9641998 (cit. on pp. 18–21).

[52] I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions", *SN Computer Science*, vol. 2, no. 3, 2021. DOI: 10.1007/s42979-021-00592-x (cit. on p. 19).

[53] S. Badillo, B. Banfai, F. Birzele, *et al.*, "An Introduction to Machine Learning", *Clinical Pharmacology & Therapeutics*, vol. 107, no. 4, pp. 871–885, 2020. DOI: 10.1002/cpt.1796 (cit. on pp. 19, 20).

[54] G. Shobha and S. Rangaswamy, "Machine Learning", in *Computational Analysis and Understanding of Natural Languages: Principles, Methods and Applications*, V. N. Gudivada and C. Rao, Eds. Elsevier, 2018, ch. 8, pp. 197–228, ISBN: 978-0-444-64042-0. DOI: 10.1016/bs.host.2018.07.004 (cit. on pp. 19, 20).

[55] M. Hossin and M. N. Sulaiman, "A Review on Evaluation Metrics for Data Classification Evaluations", *International Journal of Data Mining Knowledge Management Process*, vol. 5, no. 2, pp. 01–11, 2015. DOI: 10.5121/ijdkp.2015.5201 (cit. on p. 20).

[56] S. L. Warsof, P. Gohari, R. L. Berkowitz, and J. C. Hobbins, "The estimation of fetal weight by computer-assisted analysis", *American Journal of Obstetrics & Gynecology*, vol. 128, no. 8, pp. 881–892, 1977. DOI: 10.1016/0002-9378(77)90058-8 (cit. on p. 24).

[57] F. P. Hadlock, R. Harrist, R. S. Sharman, R. L. Deter, and S. K. Park, "Estimation of fetal weight with the use of head, body, and femur measurements - A prospective study", *American Journal of Obstetrics & Gynecology*, vol. 151, no. 3, pp. 333–337, 1985. DOI: 10.5555/uri:pii:0002937885902984 (cit. on p. 24).

[58] F. P. Hadlock, "Sonographic Estimation of Fetal Age and Weight", *Radiologic Clinics of North America*, vol. 28, no. 1, pp. 39–50, 1990. DOI: 10.1016/S0033-8389(22)01218-0 (cit. on p. 24).

[59] G. I. Hirata, A. L. Medearis, J. Horenstein, M. B. Bear, and L. D. Platt, "Ultrasonographic estimation of fetal weight in the clinically macrosomic fetus", *American Journal of Obstetrics & Gynecology*, vol. 162, no. 1, pp. 238–242, 1990. DOI: 10.1016/0002-9378(90)90857-4 (cit. on p. 24).

[60] I. Sarris, C. Ioannou, P. Chamberlain, *et al.*, "Intra- and interobserver variability in fetal ultrasound measurements", *Ultrasound in Obstetrics & Gynecology*, vol. 39, no. 3, pp. 266–273, 2012. DOI: 10.1002/uog.10082 (cit. on p. 25).

[61] H. R. Torres, P. Morais, B. Oliveira, *et al.*, "A review of image processing methods for fetal head and brain analysis in ultrasound images", *Computer Methods and Programs in Biomedicine*, vol. 215, 2022. DOI: 10.1016/j.cmpb.2022.106629 (cit. on p. 25).

[62] G. Carneiro, B. Georgescu, S. Good, and D. Comaniciu, "Detection and Measurement of Fetal Anatomies from Ultrasound Images using a Constrained Probabilistic Boosting Tree", *IEEE Transactions on Medical Imaging*, vol. 27, no. 9, pp. 1342–1355, 2008. DOI: 10.1109/TMI.2008.928917 (cit. on pp. 25, 27, 28).

[63] J. Espinoza, S. Good, E. Russell, and W. Lee, "Does the Use of Automated Fetal Biometry Improve Clinical Work Flow Efficiency?", *Journal of Ultrasound in Medicine*, vol. 32, no. 5, pp. 847–850, 2013. DOI: 10.7863/jum.2013.32.5.847 (cit. on p. 25).

[64] J. Yu, Y. Wang, P. Chen, and Y. Shen, "Fetal Abdominal Contour Extraction and Measurement in Ultrasound Images", *Ultrasound in Medicine and Biology*, vol. 34, no. 2, pp. 169–182, 2008. DOI: 10.1016/j.ultrasmedbio.2007.06.026 (cit. on pp. 25, 28).

[65] W. Wang, J. Qin, L. Zhu, D. Ni, Y.-P. Chui, and P.-A. Heng, "Detection and Measurement of Fetal Abdominal Contour in Ultrasound Images via Local Phase Information and Iterative Randomized Hough Transform", *Bio-Medical Materials and Engineering*, vol. 24, no. 1, pp. 1261–1267, 2014. DOI: 10.3233/BME-130928 (cit. on pp. 25, 28).

[66] H. Chen, D. Ni, J. Qin, *et al.*, "Standard Plane Localization in Fetal Ultrasound via Domain Transferred Deep Neural Networks", *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 5, pp. 1627–1636, 2015. DOI: 10.1109/JBHI.2015.2425041 (cit. on pp. 26, 28).

[67] H. Ravishankar, S. M. Prabhu, V. Vaidya, and N. Singhal, "Hybrid approach for automatic segmentation of fetal abdomen from ultrasound images using deep learning", in *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, IEEE, 2016, pp. 779–782. DOI: 10.1109/ISBI.2016.7493382 (cit. on pp. 26, 28).

[68] 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", *Physiological Measurement*, vol. 39, no. 10, 2018. DOI: 10.1088/1361-6579/aae255 (cit. on pp. 26, 28).

[69] X. Yang, H. Li, L. Liu, and D. Ni, "Scale-aware Auto-context-guided Fetal US Segmentation with Structured Random Forests", *BIO Integration*, vol. 1, no. 3, pp. 118–129, 2020. DOI: 10.15212/bioi-2020-0016 (cit. on pp. 26, 28).

[70] H. Gao, C. Wu, D. Huang, D. Zha, and C. Zhou, "Prediction of fetal weight based on back propagation neural network optimized by genetic algorithm", *Mathematical Biosciences and Engineering*, vol. 18, no. 4, pp. 4402–4410, 2021. DOI: 10.3934/mbe.2021222 (cit. on pp. 26, 28).

[71] M. Ferrario, M. G. Signorini, and G. Magenes, "Complexity analysis of the fetal heart rate variability: Early identification of severe intrauterine growth-restricted fetuses", *Medical & Biological Engineering & Computing*, vol. 47, no. 9, pp. 911–919, 2009. DOI: 10.1007/s11517-009-0502-8 (cit. on pp. 26, 28).

[72] J. Nithya and M. Madheswaran, "Detection of Intrauterine Growth Retardation Using Fetal Abdominal Circumference", in *2009 International Conference on Computer Technology and Development*, IEEE, 2009, pp. 371–375. DOI: 10.1109/ICCTD.2009.213 (cit. on pp. 26, 28, 30).

[73] A. V. Gadagkar and K. Shreedhara, "Features Based IUGR Diagnosis Using Variational Level Set Method and Classification Using Artificial Neural Networks", in *2014 Fifth International Conference on Signal and Image Processing*, IEEE, 2014, pp. 303–309. DOI: 10.1109/ICSIP.2014.54 (cit. on pp. 27, 28).

[74] K. s. Bagi and K. Shreedhara, "Biometric measurement and classification of IUGR using neural networks", in *2014 International Conference on Contemporary Computing and Informatics (IC3I)*, IEEE, 2014, pp. 157–161. DOI: 10.1109/IC3I.2014.7019613 (cit. on pp. 27, 28).

[75] A. Wosiak, A. Zamecznik, and K. Niewiadomska-Jarosik, "Supervised and unsupervised machine learning for improved identification of intrauterine growth restriction types", in *2016 Federated Conference on Computer Science and Information Systems (FedCSIS)*, IEEE, 2016, pp. 323–329. DOI: 10.15439/2016F515 (cit. on pp. 27, 28).

[76] M. A. F. contributors. "The Basic Steps of an Obstetric Ultrasound Examination". (2019-02), [Online]. Available: https://www.medicalaidfilms.org/film/the-basic-steps-of-an-obstetric-ultrasound-examination/ (visited on 2022-10-18) (cit. on p. 29).

[77] K. Kalyan, B. Jakhia, R. D. Lele, M. Joshi, and A. Chowdhary, "Artificial Neural Network Application in the Diagnosis of Disease Conditions with Liver Ultrasound Images", *Advances in Bioinformatics*, vol. 2014, 2014. DOI: 10.1155/2014/708279 (cit. on p. 31).

[78] M. Singh, S. Singh, and S. Gupta, "An information fusion based method for liver classification using texture analysis of ultrasound images", *Information Fusion*, vol. 19, pp. 91–96, 2014. DOI: 10.1016/j.inffus.2013.05.007 (cit. on p. 31).

[79] A. C. Bovik, "Basic Binary Image Processing", in *The Essential Guide to Image Processing*, A. Bovik, Ed. Boston: Academic Press, 2009, ch. 4, pp. 69–96, ISBN: 978-0-12-374457-9. DOI: 10.1016/B978-0-12-374457-9.00004-4 (cit. on pp. 33, 34).

[80] P. Soille, "Erosion and Dilation", in *Morphological Image Analysis: Principles and Applications*, S. Ramakrishnan, Ed. Berlin, Heidelberg: Springer, 2004, ch. 3, pp. 63–103, ISBN: 978-3-662-05088-0. DOI: 10.1007/978-3-662-05088-0_3 (cit. on p. 34).

[81] L. He, X. Ren, Q. Gao, X. Zhao, B. Yao, and Y. Chao, "The connected-component labeling problem: A review of state-of-the-art algorithms", *Pattern Recognition*, vol. 70, pp. 25–43, 2017. DOI: 10.1016/j.patcog.2017.04.018 (cit. on p. 35).

[82] R. C. Gonzalez and R. E. Woods, "Representation and Description", in *Digital Image Processing*. Upper Saddle River, NJ: Prentice Hall, 2008, ch. 11, ISBN: 0-201-18075-8 (cit. on p. 36).

[83] L. Shao and H. Zhou, "Curve Fitting with Bézier Cubics", *Graphical Models and Image Processing*, vol. 58, no. 3, pp. 223–232, 1996. DOI: 10.1006/gmip.1996.0019 (cit. on p. 38).

[84] P. Brynolfsson, D. Nilsson, T. Torheim, *et al.*, "Haralick texture features from apparent diffusion coefficient (ADC) MRI images depend on imaging and pre-processing parameters", *Scientific Reports*, vol. 7, no. 1, 2017. DOI: 10.1038/s41598-017-04151-4 (cit. on p. 39).

[85] A. K. Agwu and C. C. Ohagwu, "Histogram-based texture characterization and classification of brain tissues in non-contrast ct images of stroke patients", in *Pattern Recognition*, S. Ramakrishnan, Ed. Rijeka: IntechOpen, 2016, ch. 5, ISBN: 978-953-51-2804-5. DOI: 10.5772/65349 (cit. on p. 40).

[86] S. Naeem, A. Ali, S. Qadri, *et al.*, "Machine-learning based hybrid-feature analysis for liver cancer classification using fused (MR and CT) images", *Applied Sciences*, vol. 10, no. 9, 2020. DOI: 10.3390/app10093134 (cit. on p. 40).

[87] A. Humeau-Heurtier, "Texture Feature Extraction Methods: A Survey", *IEEE Access*, vol. 7, pp. 8975–9000, 2019. DOI: 10.1109/ACCESS.2018.2890743 (cit. on p. 40).

[88] R. Vyas, T. Kanumuri, G. Sheoran, and P. Dubey, "Co-occurrence features and neural network classification approach for iris recognition", in *2017 Fourth International Conference on Image Information Processing (ICIIP)*, IEEE, 2017, pp. 1–6. DOI: 10.1109/ICIIP.2017.8313730 (cit. on p. 41).

APPENDIX A

TEXTURAL FEATURES

The histogram features computed in the present work are explicit in Table A.1, where i denotes a grey level of the image and $P(i)$ is the first-order histogram, which is defined as follows:

$$P(i) = \frac{\text{Number of pixels with grey level i}}{\text{Total number of pixels in the image}} \tag{A.1}$$

Table A.1: Histogram Features.

Feature	Equation
Mean	$\text{Mean}(\mu) = \sum_{i=0}^{N-1} iP(i)$
Variance	$\text{Variance}(\sigma^2) = \sum_{i=0}^{N-1} (i-\mu)^2 P(i)$
Skewness	$\text{Skewness}(\mu_3) = \sigma^{-4} \sum_{i=0}^{N-1} (i-\mu)^3 P(i)$
Kurtosis	$\text{Kurtosis}(\mu_4) = \sigma^{-4} \sum_{i=0}^{N-1} (i-\mu)^4 P(i) - 3$

The GLCM texture features computed in the present work are explicit in Table A.2, where:

- N denotes the number of distinct grey levels in the image;
- $p(i,j)$ is the $(i,j)^{th}$ entry in a normalised GLCM;
- μ_i is the mean of row i in normalised GLCM;
- μ_j is the mean of column j in normalised GLCM;
- σ_i is the standard deviation of row i in normalised GLCM;
- σ_j is the standard deviation of column j in normalised GLCM.

Table A.2: GLCM Texture Features.

Feature	Equation
ASM	$\text{ASM} = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} p(i,j)^2$
Contrast	$\text{Contrast} = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} (i-j)^2$
Correlation	$\text{Correlation} = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} \frac{(i-\mu_i)(j-\mu_j)}{\sqrt{\sigma_i \sigma_j}}$
Dissimilarity	$\text{Dissimilarity} = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} p(i,j)\lvert i-j\rvert$
Energy	$\text{Energy} = \sqrt{ASM} = \sqrt{\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} p(i,j)^2}$
Homogeneity	$\text{Homogeneity} = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} \frac{p(i,j)}{1+(i-j)^2}$

2023 Semi-automatic Evaluation of Fetal Growth Restriction using Ultrasound Images Adriana Ramos