UNIVERSITÀ POLITECNICA DELLE MARCHE
FACOLTÀ DI INGEGNERIA

Master's Degree in Biomedical Engineering

Use of clustering in fetal state assessment by cardiotocography

Advisor:
Dr. Agnese Sbrollini

Candidate:
Sara Capuano

Coadvisor:
Prof. Laura Burattini

Academic Year 2023/2024

To you who have always been there,

To you who have always believed in me.

To my family.

Abstract

Hypoxia is the fetal condition that occurs when there is a reduced supply of oxygen in the fetus. Some level of hypoxia occurs in all pregnancies during childbirth, and it is due to the mechanical pressure exercised during labor. The possibility of reaching critical values for the health of the fetus depends on how and for how long time there is a reduced oxygen supply. When hypoxia reaches extreme level and it results in metabolic acidosis, it is harmful to the baby.

Nowadays, the cardiotocography (CTG) is the easiest, non-invasive, painless and widely used tool to simultaneously record the fetal heart rate signal and uterine contraction activity during pregnancy and childbirth. From its signal, by analyzing its characteristics it is possible to evaluate the fetal state assessment.

From the systematic research of studies already present in literature, emerged that a greater attention is increasingly focused on the use of artificial intelligence for clinical decision support, particularly the use of supervised ML techniques.

For this reason, this thesis has the aim to develop and implement an unsupervised ML algorithm, the CTG-GCA, able to recognize, identify and bring together cardiotocographic signals of hypoxic and non-hypoxic fetus. The proposed model was trained using the CTU-UHB database, to extract through a genetic algorithm the most relevant features of the CTG signals and subsequently, once extracted, the features were considered to create the 2 clusters using a hierarchical agglomerative clustering algorithm.

Through the model, four features were selected (BW, Area ACC, Area DEC and HF) and the statistical analysis of the clusters (Cluster 1 of 289 subjects and Cluster 2 of 263) led to the conclusion that for a correct identification of hypoxic and non-hypoxic fetus the most important features are both those strictly linked to the characteristics of the CTG signal, such as accelerations, decelerations, heart rate variability and baseline, and neonatal clinical outcome information, as the 5-minute Apgar index.

In conclusion, this work, thanks to the promising results obtained, can be considered a forerunner for this branch of research and a good starting point for future developments.

Contents

Abbreviations

ACOG	American College of Obstetricians and Gynecologists
AI	Artificial Intelligence
ANN	Artificial Neural Network
ANS	Autonomic Nervous System
AUROC	Area Under the ROC Curve
BE	Base Excess
bpm	Beat per minutes
CA	Classification Accuracy
CNN	Convolutional Neural Network
CTG	Cardiotocography
CTG-GCA	CTG - Genetic Clustering Analyzer
DNN	Deep Neural Network
DT	Decision Tree
EA	Evolutionary Algorithm
EFM	Electronic Fetal Monitoring
FHR	Fetal Heart Rate
FIGO	Federation of Gynecology and Obstetrics
GA	Genetic Algorithm
GS	Gold Standard
HC	Hierarchical Clustering
IUGR	Intrauterine Fetal Growth restriction
kNN	k-Nearest Neighbors
LSTM	Long Short-Term Memory

MANOVA	Multivariate Analysis of Variance
ML	Machine Learning
MLP	Multi-Level Perceptron
NICE	National Institute for Health and Care Excellence
O_2	Oxygen
RF	Random Forest
SVM	Support Vector Machine
UC	Uterine contraction

Introduction

Despite advances in medical field relating to the mother and baby during labour, one of the questions still open concerns the acquisition of information regarding the health of the fetus promptly to ensure opportune interventions by medical staff.

Fetal hypoxia is the condition that occurs when there is a reduced supply of oxygen to the fetus. When oxygen levels are critically low, the fetus changes its metabolism from aerobic to anaerobic, resulting in the formation of hydrogen ions in the tissues, a phenomenon known as acidosis.

Some degree of hypoxemia occurs in almost all fetus during labor due to the mechanical pressure provided by childbirth, but the possibility that it ends in hypoxia, with possible association of metabolic acidosis, depends on how and for how long time there is a reduced supply of oxygen in the fetus and its metabolic reserves at the beginning of labor. Only when hypoxia and the resulting metabolic acidosis reaches extreme levels, the fetus is at risk of long-term neurological compromise, encephalopathy hypoxic-ischemic, convulsive crisis, cerebral palsy and psychomotor delay.

In the first half of the twentieth century, fetal evaluation occurred through intermittent auscultation; only at the end of the 1960s, continuous cardiotocography (CTG) was introduced for fetal monitoring.

The CTG, from the greek *kardia* =heart and *tokos* = work, represents one of the various non-invasive fetal monitoring techniques and consists of recording the fetal heart rate (FHR, bpm) and maternal uterine activity through uterine contractions (UC, mmHg), from which it is possible to obtain the characteristics of the fetus' health that are fundamental for an accurate overview of its condition.

Despite its extensive use, cardiotocography presents some important ambiguities and discrepancies. The reasons for its limitations can be found in visual inspection and subjective interpretation of the trace, in the lack of a single and global guideline (FIGO, ACOG, NICE) and in the lack of a gold standard (GS) available that better identify and describe the newborn clinical conditions.

To overcome limits of cardiotocography, artificial intelligence is playing an important role in the field of obstetrics and gynecology as a decision support system.
This paper aims, thanks to initial systematic research, to provide clinical decision support by implementing an unsupervised machine learning algorithm capable of identifying hypoxic from non-hypoxic fetus thanks to the use of the clustering method.

Chapter 1

1. Clinical and Technical Background

1.1. Fetal state assessment

Embryogenesis and fetal growth depend on optimal maternal health and normal placenta development. Maternal exposure to a persistently hypoxic environment may lead to abnormal placenta development and negatively impact fetal growth, that can result into intrauterine fetal growth restriction (IUGR), asphyxia, multiorgan failure, premature deliver and perinatal death.
The negative impact on the fetus's health is usually referred to as "fetal distress".

1.1.1. Fetal growth

The entire human life has its roots precisely in its birth, in its development and transformation from a fertilized egg cell to a living being, totally developed. This happens through a particular state of the woman known as gestation or pregnancy.
Human pregnancy begins with the conception, or fecundation, and ends with the childbirth and it is defined as a period with a duration of 37-42 weeks during which the fetus develops [1]. It can be simple or twin if it gives rise to more than one fetus. and it can be classified also into:

- Full-term pregnancy, if the birth occurs between 37-42 weeks.
- Preterm pregnancy if the birth occurs before the 37th week.

- Proctated pregnancy if the birth occurs after the 42nd week.

Its duration can be clinically divided into three trimesters:

- The first trimester goes from the fecundation to the first 12 weeks and turns out to be the most vulnerable and precarious period.
- The second one goes from the thirteenth to the twenty-fourth week, in this phase the organs complete their development, perfecting their anatomical details.
- The last one goes from the twenty-fifth week to childbirth; some organs, such as brain, liver and kidneys, will complete their differentiation after birth but most of the others have already done so the fetus is defined as mature after exceeding 2.5 kilograms.

Unlike asexual reproduction, in which genetically identical individuals are created, in sexual reproduction this limit is exceeded. It is characterized by the fusion of the genomes of two male and female parents. The cells that contribute to sexual reproduction, the gametes, are produced by a particular form of cell division, the meiosis, and therefore they have only 23 chromosomes, half as many as any other cell. They are: spermatozoa for males and egg cells for females.
During conception, which normally occurs in the uterine tubes, contact occurs between the spermatozoon and the egg cell, giving rise to the zygote and the fusion of their chromosomal assets. The zygote is therefore the first cell of the new individual that will undergo the process of embryogenesis, forming the embryo.
From the end of the eighth week onwards all the systems are present, the embryo becomes a fetus. Is long about 3 cm, his bones started to calcify and his muscles to contract. The heart beats, circulates blood, and all the other organs and tissue start to develop. [2]

1.1.2. Fetal pathologies during pregnancy

By analyzing the development of the fetus, its organs, and tissues it is possible to notice the enormous number of steps and how the entire process can be a source of problems and complications for the fetus and the mother.

For this reason, it is important to keep them under control in order to intervene promptly if it becomes necessary to preserve the health of the fetus as well as that of the mother. Among the most common complications and pathologies it is possible to find ectopic pregnancies, spontaneous abortions, fetal deformities, but also arrhythmias and fetal hypoxia which can be more or less serious.

1.1.3. Fetal hypoxia and acidosis

During the uterine life the fetus depends on mother. Blood circulation is organized in such a way that gas exchange, the supply of nutritional materials and the elimination of catabolites take place through the placenta as the respiratory, digestive, and urinary systems are not yet able to carry out their functions. [3]
All human cells require oxygen (O_2) and its supply comes from maternal breathing and circulation, from gas exchanges across the placenta, from the umbilical and fetal circulation. If one of these systems is altered, a drop in oxygen concentration occurs at the level of the fetal arterial blood (*hypoxemia*), subsequently at the tissue level (*hypoxia*) and finally at the level of the central organs (*asphyxia*).
Fetal hypoxemia occurs in 3 ways:

1. Reduced oxygen supply to the placenta caused by:
   - hyper-stimulation or hypertonic contractions, as during labor the uterine contractions cause increases in pressure in the uterus and, consequently, on the placental vessels that supply blood to the baby. The pressure on the blood vessels therefore prevents the blood from flowing.
   - Prolonged second stage of labor.
   - Low maternal blood pressure (hypotension).
   - Maternal cardio-respiratory failure
   - Maternal hypertension
2. Reduced oxygen transfer from mother to fetus
3. Reduced oxygen transport due to fetal circulation

Hypoxia is so defined as a reduction in O_2 supply relative to the O_2 demand of the tissue. Placental oxygen varies over the course of pregnancy as O_2 delivery and metabolic demand increases with both placental and/or fetal development.
In a practical sense, the fetus in the womb experiences different types of hypoxias [4]:

1. The first type is *hypoxic hypoxia,* due to exposure to high altitude during pregnancy which results in a decrease in the amount of gases transported by the maternal blood.
2. The second type is *hypoxic asphyxia*, due to compression of the umbilical cord or interruption of the arterial blood supply to the uterine circulation, both of which occur during labor, causing both arterial hypoxemia and hypercarbia, an excessive presence of carbon dioxide in the blood.
3. The third type is *ischemic hypoxia*, resulting from disruption of perfusion of fetal tissues, such as during periods of combined reduced ventricular output secondary to conduction abnormalities or severe reflex bradycardia, which causes tissue hypoxia and hypercapnia.

Normal fetal metabolism determines the production of acids (carbonic and organic) and, despite this, the extracellular pH is maintained within a critical range thanks to the efficient buffer mechanisms of the blood system, composed of plasma bicarbonate and haemoglobin. This is because the functions of various fetal organ systems, such as the central nervous system and cardiovascular system, can be significantly affected by very small changes in pH.
When adequate fetal oxygenation does not occur, cellular energy production can still be maintained for a limited period, by anaerobic metabolism with production of organic acids, such as lactic acid, which are not easily excreted or metabolised. Accumulation of lactic acid can deplete the buffer system and cause *metabolic acidosis*.
Some degree of hypoxemia occurs in almost all fetuses during labor; the possibility that it ends in hypoxia, with possible association of metabolic acidosis, depends on how and for how long time there is a reduced supply of oxygen to the fetus and its metabolic reserves at the beginning of labor. The fetus is very resistant to hypoxia and is able to easily tolerate episodes of acidosis potentially fatal for an adult. Only when hypoxia and the resulting metabolic acidosis reach extreme levels the fetus is at risk of long-term

neurological compromise, encephalopathy hypoxic-ischemic, convulsive crisis, cerebral palsy and psychomotor delay. [4]
The evaluation of gases and lactates in umbilical cord blood or neonatal circulation during the first minutes of life represents a method for quantifying the state of hypoxia/acidosis at birth. But it is necessary and right to remember that the presence of metabolic acidosis does not exclude the presence of other factors that may cause neonatal depression and/or neurological outcomes; similarly, the absence of acidosis metabolic rate at birth does not exclude that hypoxia/acidosis may have occurred during pregnancy.

1.1.3.1. Apgar index, pH, Base Excess

The Apgar score, created by an American anaesthesiologist named Virginia Apgar, is the first test administered to the newborn, immediately after the baby is born. The test is designed to quickly assess the baby's physical condition and determine whether additional or emergency medical care is needed immediately. The Apgar score is usually assessed twice: once at 1 minute after birth and again at 5 minutes after birth, but, if necessary, also at 10 minutes after birth.
The Apgar score focuses on five main parameters, FHR, breathing, reflexes, muscle tone, colour, each of which assigns a score from 0 to 2 based on the conditions (very bad, poor, good) of the fetus; the maximum score that can be achieved it is therefore 10.
The explanation for the scoring is shown in Table 1.
Newborns with scores of 7 and above are generally normal. A score of 4-6 is considered fairly low and 3 and below are generally regarded as critically low and require a immediate medical intervention. [5]
However, the index has limitations in identifying hypoxia as it is a less than objective assessment which leaves room for interpretation, it is not altered by mild forms of hypoxia and the score can also depend on other different events.
For this reason, to obtain more concrete results regarding the health status of the fetus, Apgar it is not the only parameter taken into consideration. Because in clinical practice it cannot be quantified the oxygen concentration in the tissues, the presence of fetal hypoxia can only be ascertained from assessments of metabolic acidosis through a collection of

arterial and venous blood from the umbilical cord immediately after birth, to carry out the pH and base excess (BE) measurement.

The pH measurement is a measure of the acid-base reaction of a solution and represents the logarithm of the inverse of concentration of hydrogen ions. A pH range of 7.00 to 7.16 is cited in the literature for acidemia, but there is general consensus that a cord artery pH <7.00 is more closely related to significant with adverse neonatal outcomes and they will be more likely to suffer from short-term complications.

The BE is a blood gas parameter that is defined as the amount of acid necessary to bring the blood gas sample back to a normal pH and pHCO level at a normal body temperature and is measured in mmol/L. In other words, the basic excess is the reflection of factors that concern respiration, and which influence pH.

Newborns with BE > 8 mmol/L is recognised as moderate metabolic acidosis while severe acidosis is associated to a deficit of 12 mmol/L or greater. [5]

Promptly recognizing a risky situation is the responsibility of medical staff, as it can be associated with neonatal mortality or with the development of infantile cerebral palsy. Recognition can be carried out through the analysis of the *cardiotocographic signal*, the main tool for carrying out a correct assessment of the health status of the fetus.

	Parameter	**0 score**	**1 score**	**2 scores**
APPEARANCE	Colour	Cyanotic or pale	Cyanotic extremities	Normal
PULSE	Heart rate	Absent	< 100 bpm	> 100 bpm
GRIMANCE	Reflex stimulation	Absent	Poor	Lively, Sneezing, crying , coughing
ACTIVITY	Muscle tone	Absent	Weak flexion	Active movements
RESPIRATION	Breathing	Absent	weak or irregular	Vigours with tears

Table 1 – Apgar score assignment table

1.2. Cardiotocography

The activity of the fetal heart appears very early during pregnancy and depends on the progressive development of the nervous system of the embryo. During this process of maturation, the sympathetic and parasympathetic branches of the autonomic nervous system (ANS) play a fundamental role in controlling and modulating fetal heart rate (FHR, expressed in beats per minute - bpm).

Recording FHR and measuring its variability represent a non-invasive way to collect information about the fetal well-being and the proper development of the ANS. [6]

For these reasons, FHR monitoring is considered crucial to identify risky conditions in the fetus. Their early detection contributes to the reduction of complications and fetal death episodes as well as diminishes the need for surgical invasive interventions.

FHR can be measured using different approaches, but the most diffused is the cardiotocography (CTG).

CTG, or also known as Electronic Fetal Monitoring (EFM), was introduced in the 1960s and nowadays is the widely used tool to simultaneously record the FHR signal and uterine contraction (UC, mmHg) activity during pregnancy and childbirth. [7]

CTG is usually applied during the last trimester, thus it means after the 28th week of pregnancy, the examination is cheap, simple, non-invasive, and free of risks for fetus and mother. However, visual analysis of CTG traces requires a high level of experience on the part of obstetricians and may cause inter- and intra-observer variability.

1.2.1. Acquisition of CTG signals

Before acquiring the signal, it is necessary to consider the position to be taken by the mother that does not compromise the health of the fetus.

Changes in fetal oxygen saturation have been observed in different maternal positions. [8] The supine position, due to the aortocaval compression by the pregnant uterus, is associated with lower fetal oxygen saturation than the lateral position.

Therefore, prolonged monitoring in this position should be avoid and he lateral recumbent, half-sitting, and upright positions are preferable alternatives. [8,9]

Figure 1 – CTG device.

The CTG device provides two simultaneously recorded tracks on graph paper or electronic display: the FHR and the UC (Figure 1). The first is obtained using an ultrasound probe placed on the mother's abdomen according to the fetus's position while the second by means of a pressure transducer positioned on the bottom level of the maternal uterus as shown in Figure 2.

CTG acquisition can be performed by portable sensors that transmit signals wirelessly to a remote fetal monitor, in a telemetry modality. This solution has the advantage of allowing the mother to move freely during signal acquisition, rather than be restrained to bed or a sofa, and should therefore be the preferred option when available.

Figure 2 – Probes placement.

1.2.1.1. External vs. internal FHR monitoring

FHR detection with the external probe uses an ultrasonic Doppler transducer to detect the movement of the cardiac structures. This process results in an approximate measurement because the external monitoring is more prone to signal loss, to inadvertent monitoring of the maternal heart rate, and to signal artefacts such as double-counting and half-counting, particularly during the second stage of labor. However, it is considered an estimate sufficiently reliable.

Instead, the internal FHR monitoring using a fetal electrode (usually known as scalp electrode, but it can also be applied to the breech) evaluates the time intervals between successive heart beats by identifying R waves on the fetal electrocardiogram QRS complex. This method provides a more accurate evaluation of intervals between cardiac cycles, but it is more expensive because it requires a disposable electrode, and it requires ruptured membranes and has established contraindications, mainly related to the increased risk of vertical transmission of infections. Fetal electrode placement should also preferably be avoided in very preterm fetuses (under 32 weeks of gestation). [9]

In conclusion, external FHR monitoring is the recommended initial method for routine intrapartum monitoring, but if an acceptable record, where the basic CTG features can be recognised, cannot be obtained with external monitoring or if a cardiac arrhythmia is suspected, then internal monitoring should be used, in the absence of the contraindications.

1.2.1.2. External vs. internal monitoring of UC

External monitoring of uterine contractions using a tocodynamometer evaluates increased myometrial tension measured through the abdominal wall. This technology only provides accurate information on the frequency of contractions, and it is not possible to extract reliable information regarding the intensity and duration of contractions, nor on basal uterine tone.

Internal monitoring of uterine contractions using an intrauterine catheter provides quantitative information on the intensity and duration of contractions, as well as on basal

uterine tone, but it is more expensive as the catheter is disposable and requires ruptured membranes. [9]
As for the internal FHR monitoring, internal monitoring of UC has contraindications.
Contraindications associated with this technique include uterine haemorrhage, placenta previa, a small risk of fetal injury, uterine perforation and infection [10].
So for these reasons, internal monitoring of UC is not recommended for routine clinical use.

1.2.2. Analysis of cardiotocographic characteristics

The analysis of a CTG trace starts from the identification of the fundamental characteristics, and is followed from the overall evaluation of the CTG. Each of these assessments must be contextualized in based on the individual clinical case.
The horizontal scale for CTG registration and viewing is commonly called “paper speed” and available options are usually 1, 2, or 3 cm/min. Some experts feel that 1 cm/min provides records of sufficient detail for clinical analysis, and this has the advantage of reducing tracing length, while others feel that the small details of CTG tracings are better evaluated using higher papers speeds. In general, in the majority of the European countries is used the paper speed of 1 cm/min.
The vertical scale used for registration and viewing may also be different, and available alternatives are 20 or 30 bpm/cm.
Among the CTG characteristics is possible to find: baseline, variability, accelerations, decelerations, patterns sinusoidal and pseudosinusoidal and uterine contractions.
The following definitions are in according to the International Federation of Gynecology and Obstetrics (FIGO) guidelines.

1.2.2.1. Baseline

This is the mean level of the most horizontal and less oscillatory FHR segments. It must be estimated in time periods of 10 minutes and expressed in bpm.
It can be distinguished into [9]:

- ***Normal baseline***: if its value is between 110 and 160 bpm, but it can vary because preterm fetuses tend to have values close to the upper limit and post-term fetuses close to the lower limits.
- ***Tachycardia***: if the baseline value is above 160 bpm lasting more than 10 minutes.
  The most frequent causes of fetal tachycardia are maternal pyrexia, epidural analgesia, that may also cause a rise in maternal temperature resulting in fetal tachycardia, the administration of beta-agonist drugs, parasympathetic blockers and fetal arrhythmias, such as supraventricular tachycardia and atrial flutter.
- ***Bradycardia***: if the baseline value is below 110 bpm lasting more than 10 minutes, but values between 100 bpm and 110 bpm may occur in normal fetuses, especially in postdate pregnancies.

1.2.2.2. Variability

Variability refers to the oscillations in the FHR signal, evaluated as the average bandwidth amplitude of the signal in 1-minute segments. [9]
Also in this case is possible to distinguish:

- ***Normal variability*** with a bandwidth amplitude of 5-25 bpm.
- ***Reduced variability*** if the bandwidth amplitude is below 5 bpm for more than 50 minutes in baseline segments, or for more than 3 minutes during decelerations.
  Reduced variability can occur due to central nervous system hypoxia/acidosis and resulting decreased sympathetic and parasympathetic activity, but it can also be due to previous cerebral injury, infections or administration of central nervous system depressants or parasympathetic blockers.
  Also, during deep sleep, variability is usually in the lower range of normality, but the bandwidth amplitude is rarely under 5 bpm. During the detection, there is a high degree of subjectivity in the visual evaluation of this parameter, and therefore careful re-evaluation is recommended in borderline situations.
- ***Increased variability*** if the bandwidth value exceeding 25 bpm lasting more than 30 minutes. The pathophysiology of this pattern is incompletely understood, but it may be seen linked with recurrent decelerations, when hypoxia/acidosis evolves very

rapidly. It is presumed to be caused by fetal autonomic instability/hyperactive autonomic system.

1.2.2.3. Accelerations

It is defined acceleration an abrupt in FHR above the baseline, of more than 15 bpm in amplitude, and lasting more than 15 seconds but less than 10 minutes.
Most accelerations coincide with fetal movements and are a sign of a neurologically responsive fetus that does not have hypoxia/acidosis. Before 32 weeks of gestation, their amplitude and frequency may be lower (10 seconds and 10 bpm of amplitude). [9]

1.2.2.4. Decelerations

Decelerations are defined as a decrease in the FHR below the baseline, of more than 15 bpm in amplitude, and lasting more than 15 seconds. They can be divided into [9]:

- ***Early decelerations***, that are shallow, short-lasting, with normal variability within the deceleration and are coincident with contractions. They are believed to be caused by fetal head compression and do not indicate fetal hypoxia/acidosis.
- ***Variable decelerations (V-shaped)***, decelerations that exhibit a rapid drop, good variability within the decelerations, rapid recovery to the baseline, varying size, shape and are in relationship to uterine contractions (Figure 3).
  They are the majority of decelerations during labor and are seldom correlated to an important degree of fetal hypoxia/acidosis, until they develop worse characteristics.
- ***Late decelerations (U-shaped)*** are decelerations with a gradual onset and/or a gradual return to the baseline and/or reduced variability within the deceleration (Figure 4).

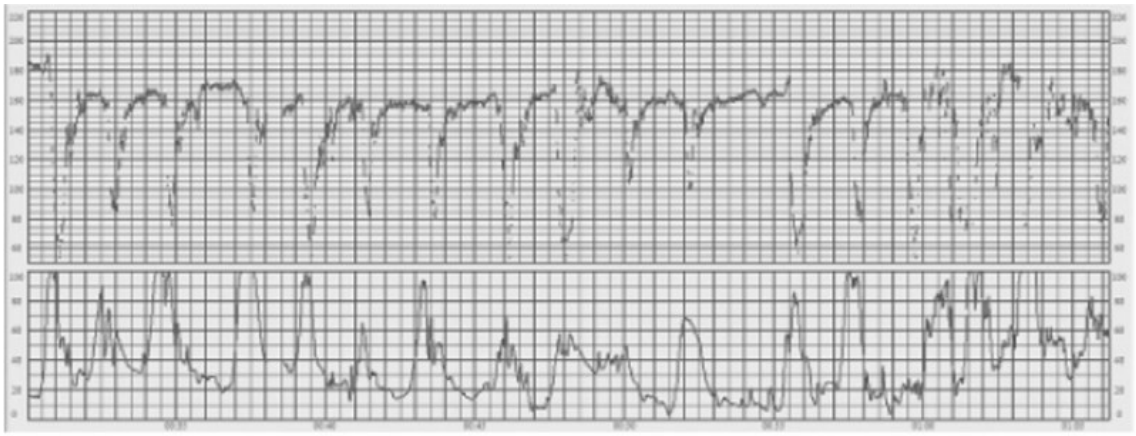

Figure 3 – Variable decelerations. Internal FHR monitoring at 1 cm/min [9]

Figure 4 - Late decelerations in the second half of the tracing. External FHR monitoring at 1 cm/min. [9]

- ***Prolonged decelerations*** are decelerations that lasting more than 3 minutes (Figure 5). These are likely to include a chemoreceptor-mediated component and thus to indicate hypoxemia. Decelerations exceeding 5 minutes, with FHR maintained at less than 80 bpm and reduced variability within the deceleration, are frequently associated with acute fetal hypoxia/acidosis and require emergent intervention.

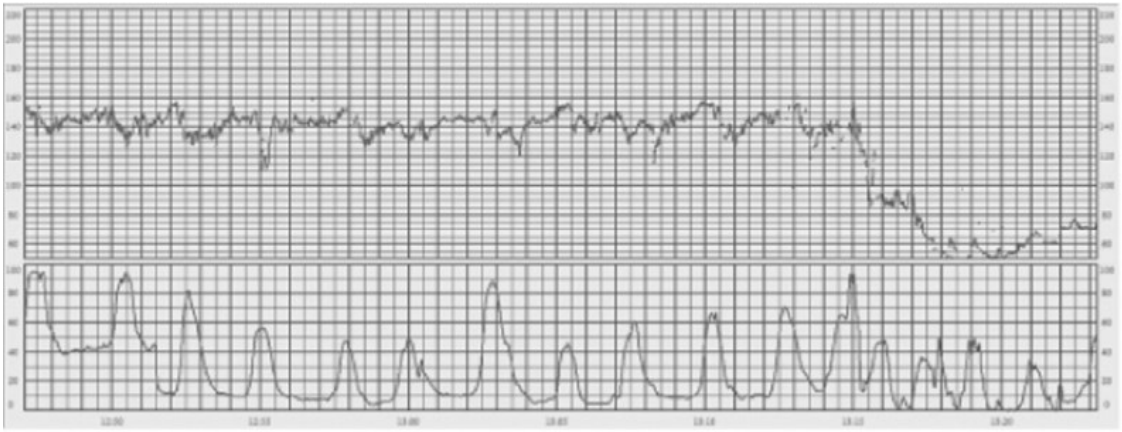

Figure 5 - Prolonged deceleration. External FHR monitoring at 1 cm/min. [9]

1.2.2.5. Sinusoidal and pseudosinusoidal patterns

The sinusoidal pattern results in a regular, smooth, undulating signal, resembling a sine wave, with an amplitude of 5−15 bpm, and a frequency of 3−5 cycles per minute. It is so defined if it lasts more than 30 minutes and coincides with absent accelerations.
The pathophysiological basis of the sinusoidal pattern is incompletely understood.
The pseudosinusoidal pattern, instead, seems to the sinusoidal pattern, but with a more jagged "saw-tooth" appearance, rather than the smooth sine-wave form. Its duration seldom exceeds 30 minutes, and it is characterized by normal patterns before and after.
It is often correlated with the analgesic administration to the mother, and during periods of fetal sucking and other mouth movements. [9]

1.2.2.6. Uterine contractions

Contractions are a critical component of the traditional definition of labor and their monitoring during labor is a widespread practice with particular attention focused on their presence, frequency, and strength. [10]
In the CTG signal, these are represented as a bell-shaped gradual increase (Figure 6) in the uterine activity signal followed by roughly symmetric decreases, with 45−120 seconds in total duration. [9]

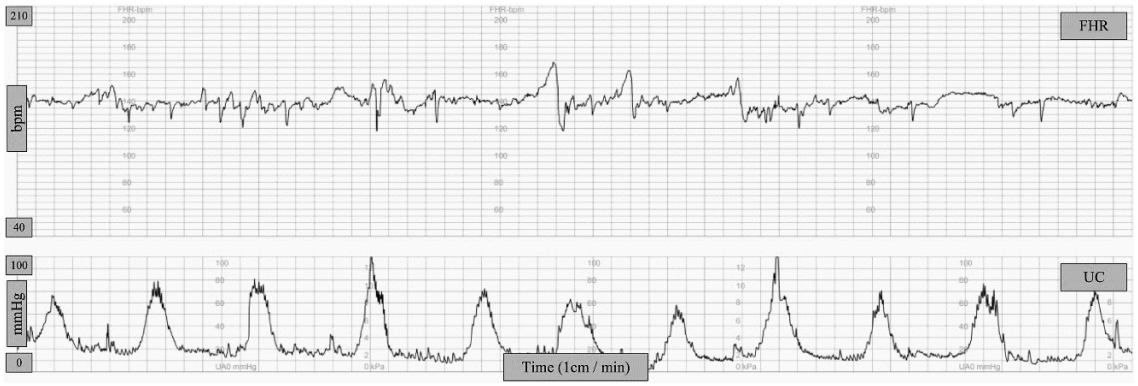

Figure 6 – Rapresentation of a tipical CTG tracing, the FHR pattern in the first figure, and the UC pattern below.

1.2.3. CTG signals classification

According to the criteria presented in Table 2 tracings should be classified into three classes: normal (type 1), suspicious (type 2) and pathological (type 3).

Due to the changing nature of CTG signals during labour, re-evaluation of the tracing should be carried out at least every 30 minutes.

In case of a normal tracing is possible to establish safely the absence of fetal hypoxia or acidosis and no action is needed to improve oxygenation of the fetus.

Hypoxia and fetal acidosis have a low chance of being found in suspicious tracing and more accurate monitoring must be performed to control the fetal oxygenation status.

In this case it is possible intervene to correct the reversible causes of hypoxia or acidosis only if they are identified.

As the name suggests, the pathological is the worst possible situation because fetus have a high probability of having hypoxia/acidosis and in this case immediate actions and additional methods are necessary to evaluate fetal oxygenation or to correct reversible causes. In acute situations immediate delivery should be accomplished

.

	Normal (type 1)	Suspicious (type 2)	Pathological (type 3)
Baseline	110-160 bpm	Lacking at least one characteristic of normality, but with no pathological features	<100 bpm
Variability	5-25 bpm		Reduced variability, increased variability, or sinusoidal pattern
Decelerations	No repetitive decelerations[a]		Repetitive[a], late or prolonged decelerations during > 30 min or 20 min if reduced variability, or one prolonged deceleration with >5 min

Table 2 - CTG classification criteria [8].
[a] decelerations are repetitive in nature when they are associated with more than 50% of uterine contractions.

1.3. Limits of cardiotocography

Despite the introduction of CTG in 1970 and its becoming a commonly used technique in the obstetric clinics, this non-invasive monitoring device presents some ambiguities and discrepancies. The reasons for its limitations can be found in visual inspection and subjective interpretation of the trace, in the lack of a single and global guideline and in the lack of a GS available that better identify and describe the newborn clinical conditions. Based on the fact that the CTG signal is analysed through visual inspection by an expert, the interpretation of the trace is subject to intraobserver and interobserver variabilities. In particular, the variation between observers (interobserved variability) refers to the condition whereby each expert may interpret the signal differently based on his skill, ability and knowledge; while the intraobserver variability is connected to the fact that the same observer, the clinician, can provide a different interpretation for the same CTG at different times.
Moreover, visual inspection suffers from poor reproducibility, error-prone and inability to recognize crucial and complex patterns in the CTG trace that can lead to misunderstandings of the CTG signal and consequently to unnecessary surgical interventions. [11]

To limit the errors due to the visual inspection of the trace and everything that derives from it, efforts in recent years have focused on the development of computer-assisted systems based on advanced machine learning (ML), with the aim of assisting obstetricians in making objective medical decision. [12]

Furthermore, the evaluation of the tracing also depends on the choosen guideline to follow, evaluate and classify the condition of the fetus. Specifically, FIGO published its first guidelines FHR monitoring in 1987, but many others national scientific organizations have also published guidelines on the subject; those with the largest impact were developed by the American College of Obstetricians and Gynecologists (ACOG) and the United Kingdom National Institute for Health and Care Excellence (NICE).These three guidelines have important differences, not only in the definition of individual CTG features but also in the criteria used for overall tracing classification, as is shown in Table 3.1, Table 3.2 and Table 3.3. [13]

Lastly, the other great limitation in the evaluation of the clinical conditions of the newborn and the fetus is precisely the lack of a GS to follow for a correct diagnosis.

Recognizing the presence of a GS has a double importance. On one hand, it may be directly used in clinical practise to optimize correct understanding of the newborn/fetus clinical status; on the other hand, it may be indirectly used as GS for the development of automatic newborn classifier, exploiting innovative ML and artificial intelligence (AI) algorithm. [14]

Feature	Guidelines		
	FIGO	**ACOG**	**NICE**
Baseline	Baseline FHR is the mean level of the FHR when this is stable, accelerations and decelerations being absent. It is determined over a time period of 5 or 10 min and expressed bpm	Mean FHR rounded to increments of 5 bpm during a 10-min segment, excluding: periodic or episodic changes, periods of marked FHR variability, segments of baseline that differ >25 bpm	Mean level of the FHR when this stable, excluding accelerations and decelerations. It is determined over a time period of 5 or 10 min and expressed in bpm
Normal baseline	110–150 bpm	110–160 bpm	110–160 bpm
Tachycardia	–	>160 bpm	>180 bpm (161–180 bpm is moderate tachycardia)
Bradycardia	<80 bpm	<110 bpm	<100 bpm (100–109 bpm is moderate bradycardia)
Variability	Oscillations of FHR around its mean level (long-term variability). This is usually only quantified by description of the amplitude of the oscillations around the baseline heart rate	Fluctuations in the baseline FHR that are irregular in amplitude and frequency. It is visually quantified as the amplitude of peak-to-trough in bpm	The minor fluctuations in baseline FHR occurring at three to five cycles per minute. It is measured by estimating the difference in bpm between the highest peak and lowest trough of fluctuation in a 1-min segment of the trace
Normal variability	Between 5 and 25 bpm	Amplitude range 6–25 bpm (moderate variability)	≥5 bpm between contractions
Reduced variability	<5 bpm for >40 min (suspicious if variability 5–10 bpm for >40 min)	Amplitude range ≤5 bpm (minimal variability)	<5 bpm for 40–90 min (non-reassuring) or >90 min (abnormal variability)
Increased variability	>25 bpm	Amplitude range >25 bpm (marked variability)	–

Table 3.1 - Different definitions of CTG's characteristics according to FIGO, ACOG, NICE guidelines [1/3]

Feature	Guidelines		
	FIGO	**ACOG**	**NICE**
Accelerations	Transient increase in heart rate of ≥15 bpm and lasting ≥15 s	A visually apparent abrupt increase (onset to peak in <30 s) in the FHR. Beyond 32 weeks of gestation, an acceleration has a peak of ≥15 bpm above the baseline, with a duration of ≥15 s but <2 min from onset to return. Prolonged accelerations last ≥2 min but <10 min	Transient increases in FHR of ≥15 bpm and lasting ≥15 s
Decelerations	Transient episodes of slowing of FHR below the baseline level of >15 bpm and lasting ≥10 s or more	–	Transient episodes of slowing of FHR below the baseline level of >15 bpm and lasting ≥15 s or more
Early decelerations	–	Visually apparent usually symmetrical gradual decrease and return of the FHR associated with a uterine contraction. A gradual decrease is defined as from the onset to the FHR nadir of ≥30 s. The decrease in FHR is calculated from the onset to the nadir of the deceleration. The nadir of the deceleration occurs at the same time as the peak of the contraction. In most cases, the onset, nadir, and recovery of the deceleration are coincident with the beginning, peak, and ending of the contraction, respectively	Uniform, repetitive, periodic slowing of FHR with onset early in the contraction and return to baseline at the end of the contraction
Late decelerations	–	Visually apparent usually symmetrical gradual decrease and return of the FHR associated with a uterine contraction. A gradual decrease is defined as from the onset to the FHR nadir of ≥30 s. The decrease in FHR is calculated from the onset to the nadir of the deceleration. The deceleration is delayed in timing, with the nadir of the deceleration occurring after the peak of the contraction. In most cases, the onset, nadir, and recovery of the deceleration occur after the beginning, peak, and ending of the contraction, respectively	Uniform, repetitive, periodic slowing of FHR with onset mid to end of the contraction and nadir >20 s after the peak of the contraction and ending after the contraction. In the presence of a non-accelerative trace with baseline variability <5 bpm, the definition would include decelerations <15 bpm

Table 3.2 – Different definitions of CTG's characteristics according to FIGO, ACOG, NICE guidelines [2/3]

Feature	Guidelines		
	FIGO	**ACOG**	**NICE**
Variable decelerations	–	Visually apparent abrupt decrease in FHR. An abrupt decrease is defined as from the onset of the deceleration to the beginning of the FHR nadir of <30 s. The decrease in FHR is calculated from the onset to the nadir of the deceleration. The decrease in FHR is ≥15 bpm, lasting ≥15 s, and <2 min in duration. When variable decelerations are associated with uterine contractions, their onset, depth, and duration commonly vary with successive uterine contractions	Variable, intermittent periodic slowing of FHR with rapid onset and recovery. Time relationships with contraction cycle are variable and they may occur in isolation. Sometimes they resemble other types of deceleration patterns in timing and shapea
Prolonged decelerations	–	Visually apparent decrease in the FHR below the baseline. Decrease in FHR from the baseline that is ≥15 bpm, lasting ≥2 min but <10 min in duration. If a deceleration lasts ≥10 min, it is a baseline change	An abrupt decrease in FHR to levels below the baseline that lasts at least 60–90 s. These decelerations become pathological if they cross two contractions (i.e. >3 min)
Sinusoidal pattern	Regular cyclic changes in the FHR baseline, such as the sine wave. The characteristics of the pattern being: the frequency is <6 cycles/min, the amplitude is at least 10 bpm and the duration should be ≥20 min	Visually apparent, smooth, sine wave-like undulating pattern in FHR baseline with a cycle frequency of three to five per minute which persists for ≥20 min	A regular oscillation of the baseline long-term variability resembling a sine wave. This smooth, undulating pattern, lasting at least 10 min, has a relatively fixed period of three to five cycles per minute and an amplitude of 5–15 bpm above and below the baseline. Baseline variability is absent

Table 3.3 - Different definitions of CTG's characteristics according to FIGO, ACOG, NICE guidelines [3/3]

Chapter 2

2. Literature review

2.1. Methods

Considering the current limitations in the evaluation of cardiotocographic tracings, in particular the inter-intra observer variability and the lack of a GS to follow, a systematic literature review was performed to evaluate the state-of-the-art of our case study.

The systematic literature search was performed according to PRISMA guidelines [15] and it was conducted in three electronic bibliographic databases: PubMed, Scopus and IEEE Xplore.

The root 'fetal*' was used to search the studies concerning the fetal topic, the roots 'CTG' and 'cardiotoco*' were used to search for studies referring to the use of the CTG trace, while the roots 'classif*' and 'cluster*'were used to focus on studies that involved the use of classification or clustering of the traces.

Terms within each concept were combined with the Boolean operator 'OR' and then combined with the Boolean operator 'AND'. As a limit for the field of search, 'Title/Abstract' was used for all the concepts, expect for the roots 'CTG' and 'cardiotoco*' where 'Title' was used.

Moreover, English language and open access were used as limits to filter the documents.

The outcomes of the literature search were imported in Zotero reference management system to remove duplicates. Documents that were not available were excluded as well as those documents like reviews or databases were removed.

Then, titles and abstracts were examined to select only documents of interest and finally, a full text analysis was carried out to discard documents according to following these exclusion criteria:

1. The aim of the study was to identify hypoxic from non-hypoxic fetuses;
2. The database used was the CTU-UHB.

2.2. Results

As shown the flow-chart of the entire process of literature search and study selection (Figure 7) the search resulted in 125 documents: 37 from PubMed, 57 from Scopus and 31 from IEEE Xplore. Of these, 31 were duplicated and therefore removed.
Moreover, analysing documents according to title and abstract, 56 were removed obtaining 34 articles and finally, performing a full-text examination based on the chosen exclusion criteria, 5 articles were selected [12, 16-20].
In the following sections is given a general overview of the methods proposed in literature for the evaluation of the clinical assessment of the fetus through clustering techniques.

2.2.1. Ajirak et al.

In this paper [12] is studied the classification of the CTG signals into hypoxic and normal classes. The main challenge of the study lies in the highly imbalanced data which will lead to poor performance of the ML methods.
To address this issue is adopted a method known as boot ensemble learning, that iteratively allows ML methods to retrain by using in every iteration more informative data samples.
After the feature selection and the elimination of the redundant ones, the selected ones were used in the following algorithms: Random Forest (RF), Naïve Bayes, AdaBoost, k-Nearest Neighbors (kNN), Support Vector Machine (SVM), and Decision Tree (DT).
A detailed comparison of these methods using different balancing approaches on CTG signals available from a public database has been released (Table 4).

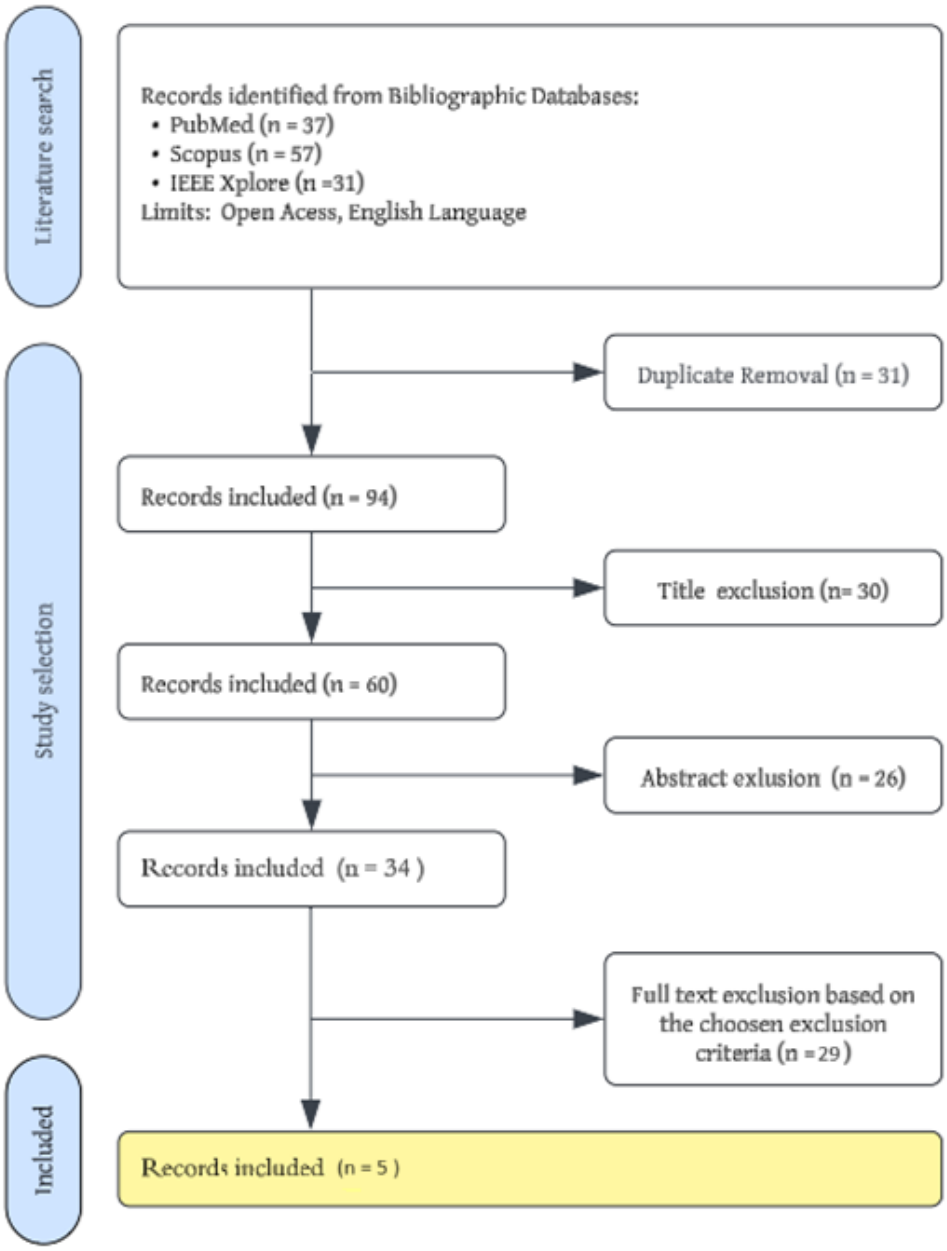

Figure 7 - Process of literature search and study selection.

As result has been obtained that the proposed boost ensemble undersampling clearly shows the best performance for all the applied methods, in terms of Area Under the ROC Curve (AUROC), Classification Accuracy (CA), F1 score, Precision, and Recall.

	AUROC	CA	F1	Precision	Recall
No resampling					
SVM	0.684	0.837	0.374	0.486	0.304
Random forest	0.753	0.831	0.033	0.200	0.018
Decision Tree	0.626	0.819	0.452	0.441	0.464
Ada Boost	0.657	0.788	0.413	0.371	0.464
kNN	0.477	0.650	0.164	0.133	0.214
Boost ensemble undersampling					
AdaBoost	0.796	0.798	0.785	0.824	0.750
Decision Tree	0.799	0.798	0.789	0.811	0.768
Random forest	0.887	0.789	0.774	0.820	0.732
SVM	0.826	0.772	0.750	0.812	0.696
kNN	0.494	0.518	0.574	0.507	0.661

Table 4 – Comparison of resampling methods

2.2.2. Daydulo et al.

The key contribution of the present study [16] consists in preprocessing of the raw FHR signal to remove unwanted artifacts and missing data, in the conversion of 1D FHR signal to 2D image by application effective time-frequency representation method and finally in the development of ResNet50, a deep learning model for training and classification into normal and distress cases of the first 20 min (experiment 1) and the last 15 min of CTG recording (experiment 2). This process is reported in Figure 8.

The study used the transfer learning to avoid the building and the training of a network from scratch. It permits the use of model that have already been trained on huge dataset to train models with less labelled data and in this specific case the convolutional neural network (CNN) used is the ResNet.

ResNet was introduced in 2015 and its structure (Figure 9) makes it one of the most powerful deep neural networks and examining the performance for both experiments it achieves an accuracy of 98,7%, sensitivity of 97,0% and specificity of 100% for the first experiment and an accuracy of 96,1%, sensitivity of 94,1% and specificity of 97,7% for the 2nd last stage of recording using CTU-UHB dataset.

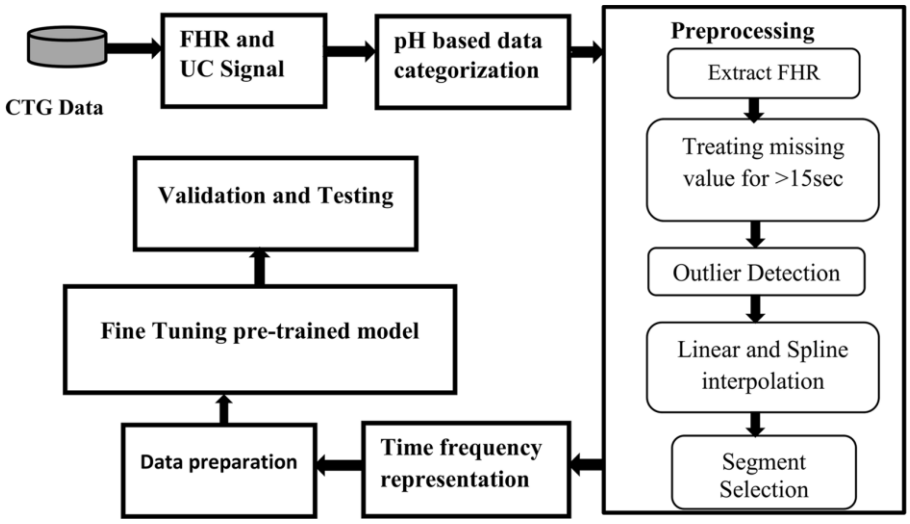

Figure 8 – Scheme of the overall methodology [16]

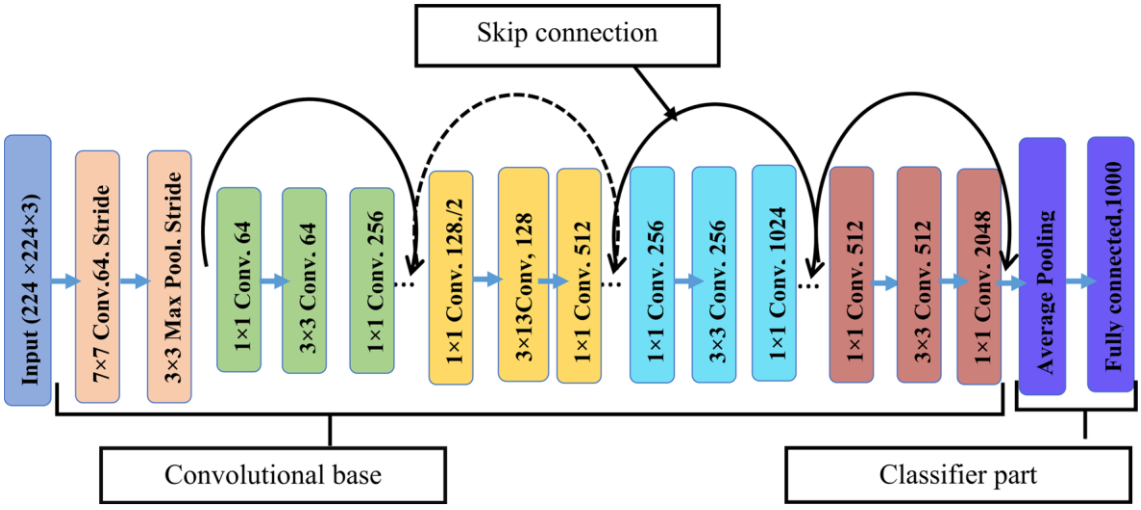

Figure 9 – Architecture of ResNet50 [16]

2.2.3. Ogasawara et al.

The study [17] address the challenge of directly predicting infants' perinatal outcomes from CTG pattern by deep Neural network (DNN) based approach. DNN is a method of

ML composed of multiple layers that automatically extract hierarchical features, similar to the human mechanism.

The architectures adopted in the study are CNN, and in particular it is used the CTG-net, and long short-term memory (LSTM) whose performances are then compared with the conventional machine learning techniques: SVM and k-mean clustering.

In the study it was demonstrated that DNN models could predict infant outcomes and looking to the performances (F1 score and AUROC) were significantly better than those of conventional algorithms. In addition, it is clarified that the CTG-net model achieved significantly higher performance than the LSTM model (AUC of CTG-net was 0.71 ± 0.04 compared to the AUC of LSTM model of 0.61 ± 0.02).

2.2.4. Francis et al. (2022)

The study [18] proposed a comparison between 5 ML classifier, which include DT, RF, SVM, kNN, SVM and Artificial Neural Network (ANN).

From the results of the study, it emerged that the RF classifier has the highest F1 score (97.00%), AUROC (99.73%) and precision (97.00%), while the other results from different classifiers are all recorded in Table 5.

Compared with other studies, the performances measure in this search are as high as those that used pH levels as a surrogate marker for hypoxia. This indicates that Apgar scores are as good as pH levels in classifying hypoxia for the CTU-UHB dataset.

Classifier	Precision (%)	Recall (%)	F1 score (%)	AUROC (%)
DT	93.0	88.0	90.0	94.9
RF	**98.0**	**98.0**	**98.0**	**99.8**
SVM	72.0	71.0	71.0	76.9
kNN	70.0	85.0	77.0	81.1
ANN	51.0	100.0	67.0	50.0

Table 5 – Comparison of performance metrics between classifiers. [18]

2.2.5. Francis et al. (2023)

If from the previous study emerged that the Apgar index can also be used as a benchmark for hypoxia, in this paper [19], five different low Apgar score boundaries has been tested to investigate how changes in the boundaries affect the performance metrics in classifying abnormal CTG.

The classifiers tested are RF and multi-level perceptron (MLP). CTG signal was also classified using pH (pH < 7.05) as a benchmark to compare the result between Apgar score and pH for classifying hypoxia.

From the results reported in Table 6 it is possible to find that the Apgar score defined as less than 10 showed higher performances in F1, precision and recall; while the highest AUROC was obtained when the low Apgar score is defined as less than 9.

So, it seems ideal to set the low Apgar score of less than 10 in terms of prediction performance, but in clinical practise, however, scores between 7 to 10 are considered healthy, meaning that scores 7, 8 and 9 will be misclassified as hypoxic, increasing the number of cases of false positive.

Buondary for low	Model	F1 score (%)	Precision (%)	Recall (%)	AUROC (%)
Apgar <10	RF	63.28	68.31	59.33	63.21
	MLP	59.61	61.68	57.78	56.31
Apgar < 9	RF	41.84	36.7	49.17	68.12
	MLP	34.37	35.93	33.93	60.47
Apgar < 8	RF	10.62	5.29	11.11	50.23
	MLP	15.41	12.22	8.89	52.91
Apgar < 7	RF	15.24	6.23	27.78	56.99
	MLP	10.52	2.56	5.56	52.39
pH < 7.05	RF	62.31	56.75	58.33	76.18
	MLP	34.43	35.50	34.65	75.09

Table 6- Performances of RF and MLP with different boundaries of Apgar score and comparison with pH score. [19]

2.3. Comparison table and discussion

Table 7 provides a comparison of the main features of the studies analyzed in this systematic literature review and from it is possible to deduce that all the studies taken into consideration use the CTU-UHB as database and are focused on the classification of subjects as hypoxic or not.

Furthermore, as could be expected, the lack of a GS to follow for the correct classification and identification of the status of fetus is highlighted. In fact, not only there is a disagreement on the neonatal clinical outcomes to take into consideration as a GS (pH or Apgar), but also on its threshold value.

Below, Figure 10 contains the frequency with which the classification algorithms were used in the studies considered. The most frequently analyzed ones are the supervised algorithms: RF and SVM.

As regards the metrics for evaluating the performance of the algorithms, all the studies considered suggest the use of the classic evaluation methods of supervised algorithms.

These measures based on predictive accuracy compared to already known labels, include accuracy, precision, recall, F1-score and AUC-ROC Curve.

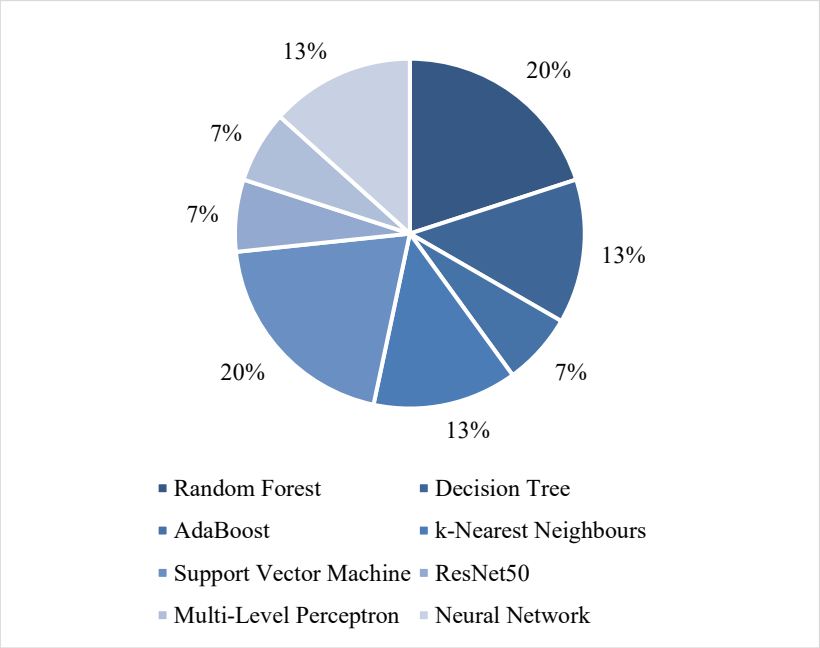

Figure 10 – Frequency at which the algorithms are used in the considered studie.

Ref.	Title	Author	Year	Database	Classification	Gold Standard	Methods used	Evaluation Metrics
[12]	Boost Ensemble Learning for Classification of CTG SIGNALS	Ajirak M, Heiselman C, Quirk JG, Djurić et al.	2022	CTU-UHB	Hypoxic / Not Hypoxic	pH < 7.05	Random Forest (RF), AdaBoost, k-Nearest Neighbours (kNN), Decision Tree (DT), Support Vector Machine (SVM)	AUROC, Accuracy, F1 Score, Precision, Recall
[16]	Deep learning based fetal distress detection from time frequency representation of cardiotocogram signal using Morse wavelet: research study	Daydulo, Y.D., Thamineni, B.L., Dasari, H.K. Et al.	2022	CTU-UHB	Normal / Distress cases	pH < 7.15	ResNet50	Accuracy, Sensitivity, Specificity
[17]	Deep neural network-based classification of cardiotocograms outperformed conventional algorithms	Ogasawara J, Ikenoue S, Yamamoto H, et al.	2021	KEIO University hospital + CTU-UHB	Normal/ Distress cases	Apgar1 < 7 OR pH < 7.20	Convolutional Neural network (CNN) , LSTM compared with SVM and k-mean clustering	AUROC, F1 Score
[18]	Detecting Intrapartum Fetal Hypoxia from Cardiotocography Using Machine Learning	F. Francis, H. Wu, S. Luz et al.	2022	CTU-UHB	Hypoxic / Not Hypoxic	Apgar < 7	DT, RF, SVM, kNN, Artificial neural network (ANN)	Precision, Recall, F1 score, AUROC
[19]	Machine Learning to Classify Cardiotocography for Fetal Hypoxia Detection	F. Francis, H. Wu, S. Luz et al.	2023	CTU-UHB	Hypoxic / Not Hypoxic	Apgar5 < 10 / 9 / 8 / 7	RF, Multi-Level Perceptron (MLP)	F1 score, Precision, Recall, AUROC

Table 7 – Comparison among different studies

Chapter 3

3. Use of clustering for fetal state assessment

3.1. Materials and Methods

3.1.1. Dataset

The data used in this research for the develop of a clustering method for the classification of hypoxic or not hypoxic fetus come from an open-access intrapartum CTG database described in [20], freely available on PhysioNet [21]. The database, from the Czech Technical University (CTU) in Prague and the University Hospital in Brno (UHB), contains 552 CTG recordings, which were carefully selected from 9164 recordings collected between 2010 and 2012 at UHB.

Each CTG contains the FHR time series and the UC signal, each sampled at 4 Hz of various length acquired from the mother during the last phase of the labor, immediately before delivery.

Each recording is characterized by 57 features, including clinical maternal information, neonatal clinical outcomes, FHR, UC and frequency information as reported in Table 8.1 and Table 8.2.

Furthermore, in the present study, missing data were replaced with the number 0.

Feature	Definition	Unit of measure
Clinical maternal information		
GAge	Week of gestation	Weeks
MumAge	Womens's age	Years
Diabetes	Presence of diabetes	n.a.
Gravidity	Total number of pregnancies	n.a.
Parity	Number of delivery completed	n.a.
Hypertension	Presence of hypertension	n.a.
Preeclampsia	Presence of preeclampsia	n.a.
Pyrexia	Presence of fever	n.a.
Liq. Praecox	Presence of early amniotic fluid	n.a.
Meconium	Presence of meconium in amniotic fluid	n.a.
Neonatal clinical outcome information		
BW	Birth weight	Gram (g)
pH	pH value in blood	pH unit
Bdecf	Decelerative flow of the fetal heartbeat	mmol/L
pCO2	Partial pressure of carbon dioxide in the blood	mmHg
BE	Base Excess	mmol/L
Apgar1	Apgar score 1 minute after birth	n.a.
Apgar5	Apgar score 5 minutes after birth	n.a.
FHR informations		
#ACC	Number of ACC during labor	n.a.
Dur ACC	Average ACC duration	Seconds (s)
Amp ACC begin	Average amplitude of the begin of ACC	bpm
Amp ACC end	Average amplitude of the end of ACC	bpm
Amp ACC middle	Average amplitude in the middle of ACC	bpm
Area ACC	Average area under the ACC curve	bpm*s
#DEC	Number of DEC during labor	n.a.
Dur DEC	Average DEC duration	s
Amp DEC begin	Average amplitude of the begin of DEC	bpm
Amp DEC end	Average amplitude of the end of DEC	bpm
Amp DEC middle	Average amplitude in the middle of the DEC	bpm
Area DEC	Average area under the DEC curve	bpm*s
#Early	Number of early DEC	n.a.
#Variable	Number of variable DEC	n.a.
#Late	Number of late DEC	n.a.
#Prolonged	Number of prolonged DEC	n.a.
LTVI	Index of long-term variability	ms
LTV6	Long-term variability at 6 minutes	ms
LTV25	Long-term variability at 25 minutes	ms
ALTV	Absolute long-term variability	ms
LTV	Long-term variability	ms
STV	Short-term variability	ms
SDSTV	Standard deviation of short-term variability	ms
STV1	First value of short-term variability	ms
STVI	Last value of short-term variability	ms
BL range	Baseline range	bpm
BL 110	Baseline at 110bpm	s
BL 160	Baseline at 160bpm	s
Mean BL	Mean Baseline	bpm

Table 8.1 – Definition of the CTU-UHB features [1/2]

Features	Definitions	Unit of measure
UC informations		
#UC	Number of UC during labor	n.a.
Dur UC	Average UC duration	s
AmpUC begin	Average amplitude of the begin of UC	mmHg
Amp UC end	Average amplitude of the end of UC	mmHg
Amp UC middle	Average amplitude in the middle of UC	mmHg
Area UC	Average amplitude under the UC curve	mmHg*s
Frequency information		
VLF	Variance low-frequency	Hz
LF	Low-frequency component of the FHR	Hz
MF	Average frequency	Hz
HF	High-frequency component of the FHR	Hz
LF/(MF+HF)	Ratio between LF and the sum of MF and HF	n.a.

Table 8.2 – Definition of the CTU-UHB features [2/2]

3.1.2. The CTG–Genetic Clustering Analyzer (CTG-GCA)

The CTG-Genetic Clustering Analyzer (CTG-GCA), described in this section, is an unsupervised ML method to extract the main features form CTG signals that better characterize a hypoxic fetus from a non-hypoxic one.

ML refers to that part of AI that it is focused on setting machines to be able to perform some tasks without the need of explicit programming, but with the ability to learn by itself. Based on the available data it is possible to distinguish 3 types of ML: supervised, unsupervised, and reinforced ML techniques.

Unlike the supervised model, where the date and class label are provided, in the unsupervised models the learning system is supposed to detect patterns without pre-existing labels or specifications. Therefore, the training data consists only of variables with the goal of finding structural information of interest [25,26].

Specifically, the CTG-GCA developed in this study and depicted in Figure 11, takes as input the list of features without any target references, and proposes two steps implementation of two different algorithms.

In the first step, with a feature selection algorithm the most relevant features for our problem of interest were extracted.

Subsequently, in a second step, the most important features resulting from the extraction were considered for the partition of the non-labeled data into a certain number of clusters.

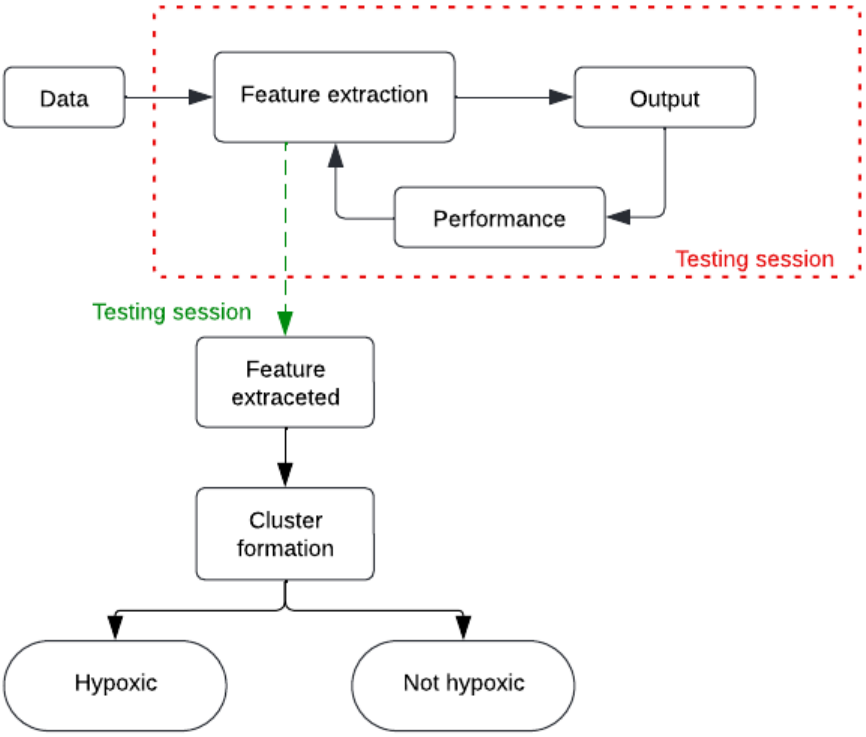

Figure 11 – Architecture of the unsupervised CTG – GCA model.

In particular, in this study the clusters formed are related to hypoxic and non-hypoxic fetuses.

3.1.2.1. Genetic Algorithm

Darwinian evolution of 1895 was used as starting point in introducing Evolutionary Algorithms (EAs) made up of techniques that transpose into algorithms of optimization the principles of natural evolution.

The key aspect that distinguishes EAs from traditional methods is that it is population-based, on which the algorithm evolves generation after generation, performing a search for the solution efficiently. The most popular evolutionary computing technique is the Genetic Algorithm (GA). [22]

GAs, described for the first time by Holland in 1975, are heuristic methods based on Darwin's principle "Survival of the fittest".

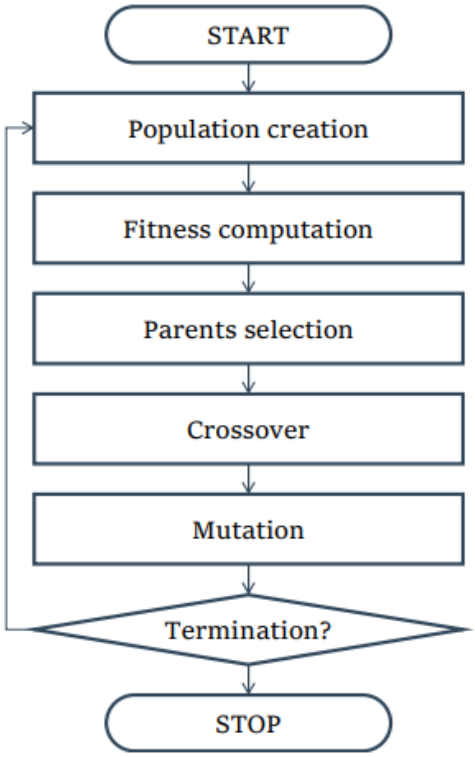

Figure 12 – Flow diagram of GAs

From the architecture of the GAs, represented in Figure 12, it emerges that the iterative process is the key point of the algorithm, which involves the manipulation of a population. The *population*, in a GA, is a set of individuals, or called *chromosomes*, and each chromosome is made up of a certain number of *genes,* which in binary representation are represented by 0 and 1 (Figure 13).

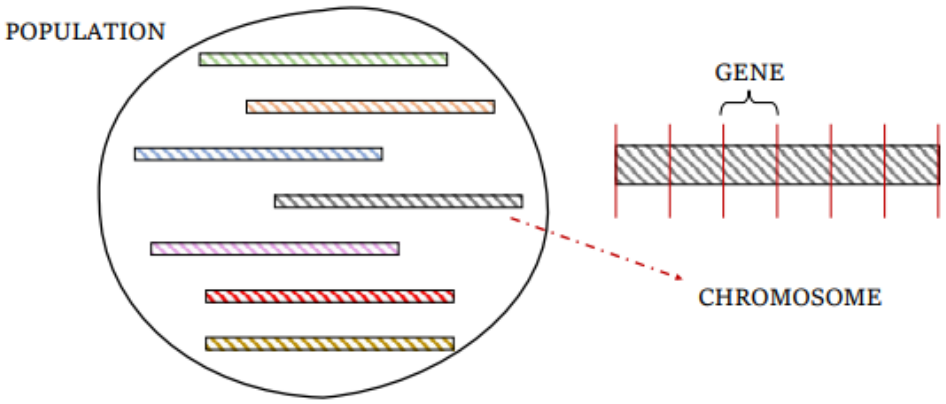

Figure 13 – Representation of population, chromosomes and genes

Manipulation involves the creation of a new population made up of new and possibly more suitable individuals, which are recombined so that they evolve towards a maximum point. The function to be maximized is called *fitness* and until the stopping criteria are not satisfied the algorithm consists into: [23-24]

- *Parents selection*, that is, the selection of individuals (parents) whose genetic makeup it must be inherited, in whole or in part, by the subsequent population. The selection process can take place based on the highest fitness value, or randomly if the fitness of the individuals is similar (Figure 14).
- *Crossover*, the process through the which the parents' genes are mixed to give rise to different *offspring*. There are different methodologies of crossover, based on a point, point to point or multipoint crossing and its frequency depends on the applications. It can be applied all the times, in the way in which all offspring are made via crossover, only in some generations, or never, creating offspring equal to the parents (Figure 14). [23]
- *Mutation,* that provides a small element of randomness in the individuals of the populations. In this case the mutation frequency depends on the application and it can be applied never (0%), so that all the genes of offspring are equal to genes of parents, rarely (≈1%) adding some diversity, often (≈50%) applying a huge variability and always, such that all the genes of offspring are different (Figure 14).

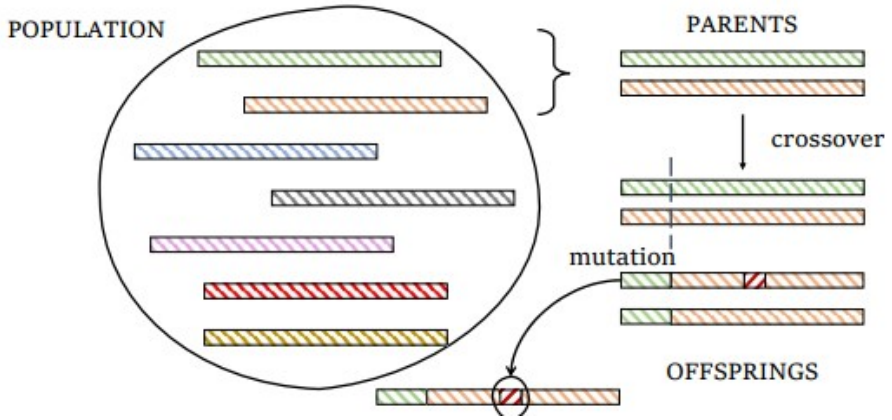

Figure 14 – Representation of parents selection, crossover and mutation

- *Offspring replacement* is the last stage of the iterative process of GAs in which the offspring are inserted into population.

The crucial point of the whole algorithm is the fitness computation, the definition of which depends on the singular application. However, as regards the stopping criteria, it is possible to use different criteria, such as the achievement of the maximal number of generations, of some level of fitness, of the minimal number of diversity or of the number of generations with no significant change of fitness.

In summary, GA represent a useful tool for search and optimization problems and its advantages are ease of coding, ease of modification according to applications, can be parallelized, can provide multiple solutions, it is resistant to becoming trapped in local optima, it performs very well for large-scale optimization problem and it can be employed for a wide variety of optimization problems.

Among the limitations, however, there is difficulty of identifying fitness function, it is slow, and premature convergence can occur [22].

3.1.2.2. Dendrogram

Clustering is an example of unsupervised ML where no labelled data are available. The goal of clustering analysis is to partition a set of object or data into a number of more or less homogenous subgroups or clusters on the basis of a measure of similarity.

The creation of clusters is done in such a way that the similarity between objects within a cluster is larger than the similarity between objects belonging to different subgroups. Since similarities are difficult to find, as similarity is a subjective measure, it may be easier to look at differences, distances, or dissimilarities between objects of the same or different groups. [27]

According to the methodologies and reasoning behind the creation of the clusters, two different cluster algorithm types exist: the partitional and the hierarchical algorithms.

If the partitional ones consist in the division of a set of n objects, in a k number of clusters ($k<n$) and in the iterative reallocation of the objects; the hierarchical (HC) ones instead involve the hierarchical decomposition of the data and based on this it is possible to distinguish them into divisive ones or agglomerative.

Agglomerative clustering starts with clusters, including each of them exactly one object, and a series of merge operations are then followed out that finally lead all objects to the same group. Divisive clustering proceeds in an opposite way. In the beginning, the entire data set belongs to a cluster and a procedure successively divides it until all clusters are singleton clusters. [28]

The results of both hierarchical algorithms can be depicted by a dendrogram, a tree-like structure represented in Figure 15, where the root nodes, on the x-axis, represent the whole dataset and each leaf node is regarded as a data object. The intermediate nodes, thus, describe the extent that the objects are proximal to each other; and the height of the dendrogram, on y-axis, usually expresses the distance between each pair of objects or clusters, or an object and a cluster.

Since in hierarchical clustering, while constructing the dendrogram, there is no need to define the number of clusters in advance; the ultimate clustering results can be obtained by cutting the structure horizontally (red lines in Figure 15) at different levels.

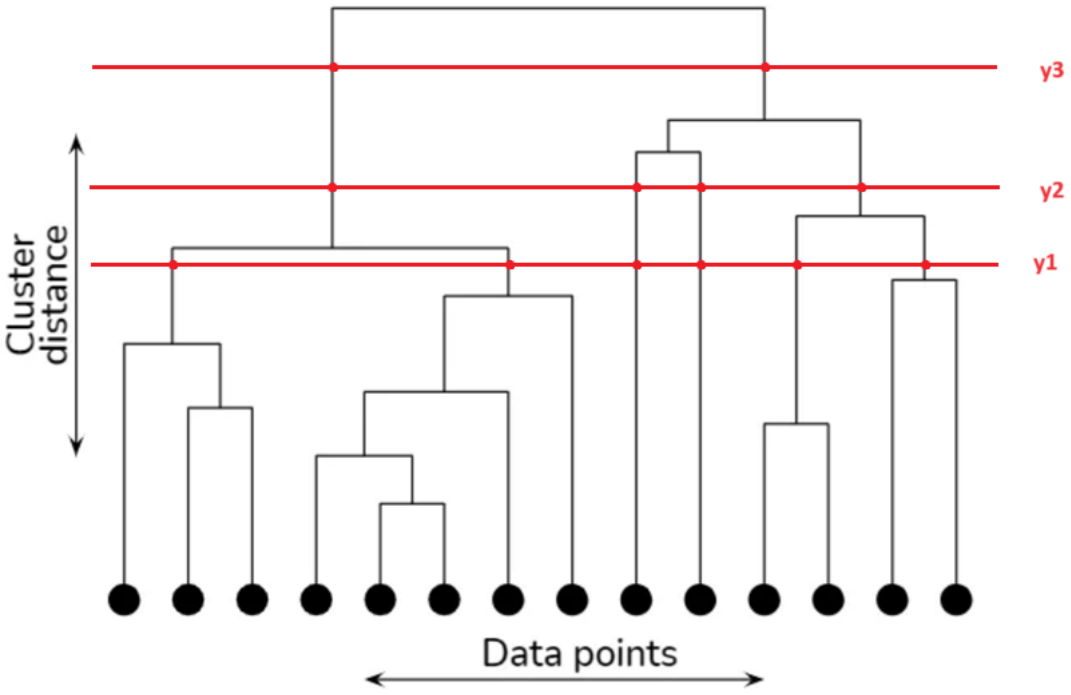

Figure 15 – Dendrogram with data points on x-axis and cluster distance on y-axis.

In conclusion, if on one hand is possible to define the dendrogram as a very intuitive and graphical method to solve clustering problems, on the other hand dendrograms, like hierarchical algorithms, suffer from lack of robustness and therefore are sensitive to noise and outliers. As a result, once an object is assigned to a cluster, it will not be considered again, meaning that HC algorithms are unable to correct any previous classification errors.

Furthermore, dendrograms, although they seem mathematically simple to understand, are computationally heavy as the calculation of the distances between the samples and the subclusters significantly increases the number of operations to be performed.

Moreover, as can be seen from its representation, the method being graphical has the disadvantage of becoming impossible to read, study and understand if the number of data samples increases.

3.1.3. Statistical analysis

Since the proposed model, the CTG-GCA, is an unsupervised model its evaluation could not be based on the predictive accuracy with respect to the already known labels, but on measures of internal consistency between data and separation between clusters.

In this study, the proposed evaluation metric is the Silhouette index.

The Silhouette coefficient is an index able to evaluate the quality of clusters of points in terms of cohesion and separation.

To calculate the Silhouette index is necessary know the clusters, already obtained by the application some clustering techniques, and the collection of all distances between objects. In particular, 2 average distances must be computed:

- *Cohesion a(i),* defined as the average distance between the point in question *i* and all the other points belonging to the same cluster, assuming that it is cluster A.
- *Separation b(i)*, defined as the average distance between the point *i* and all the points of the closest cluster other than that of belonging, cluster B.

So, at this point, for each object *i* a certain value of Silhouette *s(i)* is obtained [29]:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i),\ b(i)\}} \qquad \text{Eq. 1}$$

And as shown in Eq. 1, *s(i)* is defined as the ratio between the difference between the average distance of the nearest cluster and the average distance from the own cluster and the maximum value between the two distances, in order to normalize the result.
The Silhouette value is therefore in the range between [-1, 1].
Values close to 1 indicate that the point is well grouped within its own cluster and well separated from other clusters. Values close to 0 indicate that the point is very close to the boundary between two clusters and negative values indicate that the point is probably assigned to the wrong cluster.
For this reason, the silhouette index is a measure to evaluate the quality of the clusters created by quantifying how well each point is grouped within its cluster compared to the others.

3.1.4. Implementation

The entire model was implemented on MATLAB ® R2024a environment.
The dataset was divided into two sets: training set which contained 70% of the data that were used to train the implemented ML method and the test set containing the last 30% of the data that were kept away from the system during training and were used only to evaluate it. The described data split is shown in Figure 16.
The CTG -GCA, as has already been described in the paragraph 3.1.2, is based on the use of GA for feature extraction and the creation of dendrograms with regards to the creation of data clusters based on the selected features.
In particular, the implementation characteristics of the GA are the following:

- Number of individuals in the population: 50.
- Population initialization: randomization of 0 and 1.
- Length of the chromosomes: 57.
- Number of parents selected based on the highest fitness level: 4.
- Number of offspring generated at each iteration: 4.
- Mutation rate: 60%.
- Crossover point applied casually.
- Replacement of offspring through fitness comparison.

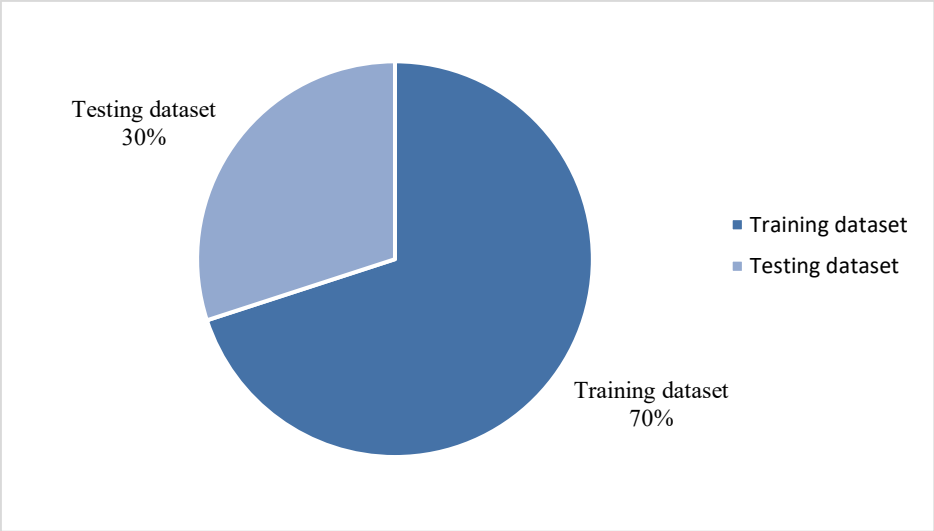

Figure 16 – Dataset division.

- Stopping criteria: steady-state fitness for 100 generations or fitness level achievement equal to 1.

While as regards the definition of the two clusters (hypoxic or non-hypoxic), firstly the distances between the fetus of the CTU-UHB database were computed and subsequently hierarchical agglomerative clustering was performed using Ward's linkage criterion on the calculated distances, minimizing the variance within the clusters.

The feature extraction was firstly implemented in a supervised manner, considering pH as the GS (with an identification threshold between healthy and diseased of 7.10 [14]) and not considering the neonatal clinical outcomes (pH, Apgar1, Apgar5, BE) and subsequently in an unsupervised manner. In both cases 100 feature extraction trials were carried out.

If in the supervised method the GA fitness function was designed to calculate the fitness of a logistic regression model that measures the AUROC curve, in the unsupervised one two clusters were defined within the fitness function and for their evaluation the silhouette index was evaluated.

Lastly the clusters thus defined, considering the extracted features in the previous steps, were evaluated from a statistical point of view and furthermore the average and standard deviation of each feature was calculated for both clusters. The analysis consists of comparing the data from each group to determine if there are significant differences

between them. This process often involves the use of statistical tests that produce p-values. Depending on the value of the p-value it is possible to say whether there is a statistically significant difference between the groups or clusters (p-value < 0.05) or not (p-value ≥ 0.05).

The results of the study are listed in the following paragraph.

3.2. Results

The results of the study are reported below.

Histograms depicted in Figures 17 and 18 represent the features that were extracted using the proposed model after carrying out the 100 trials.

In all graphs, the x axis shows the list of features (number of features = 53), excluding clinical neonatal outcomes, while the y axis shows the frequency of selection.

Figure 17 represents the frequency with which the features were selected by implementing the supervised model which takes pH into consideration as the gold standard.

Figures 18 instead show the features that were selected with the CTG-GCA model through the Silhouette index evaluation.

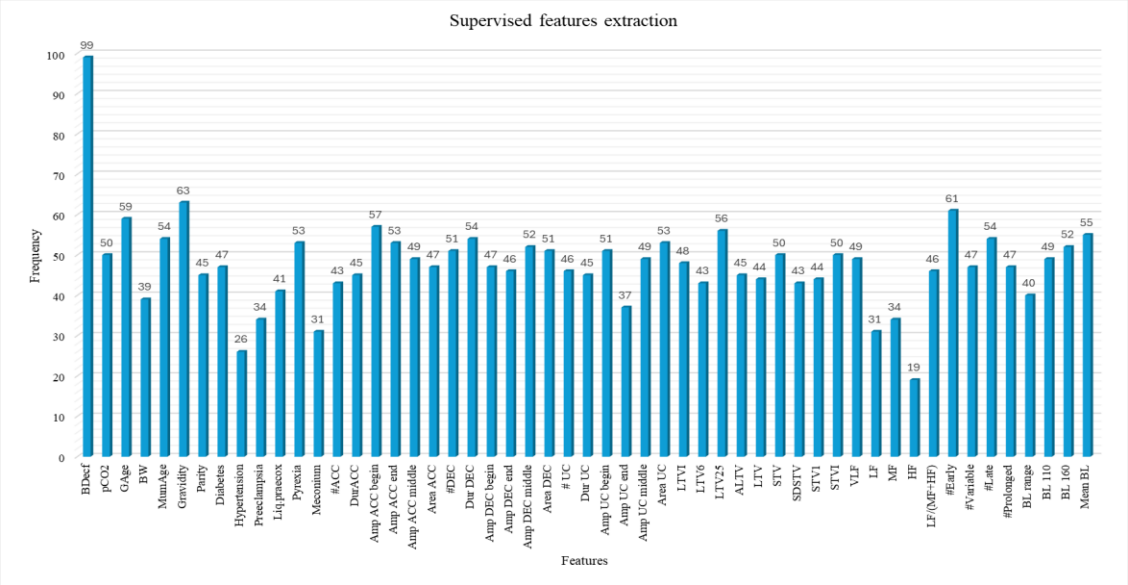

Figure 17 – Features extraction frequency via supervised ML method

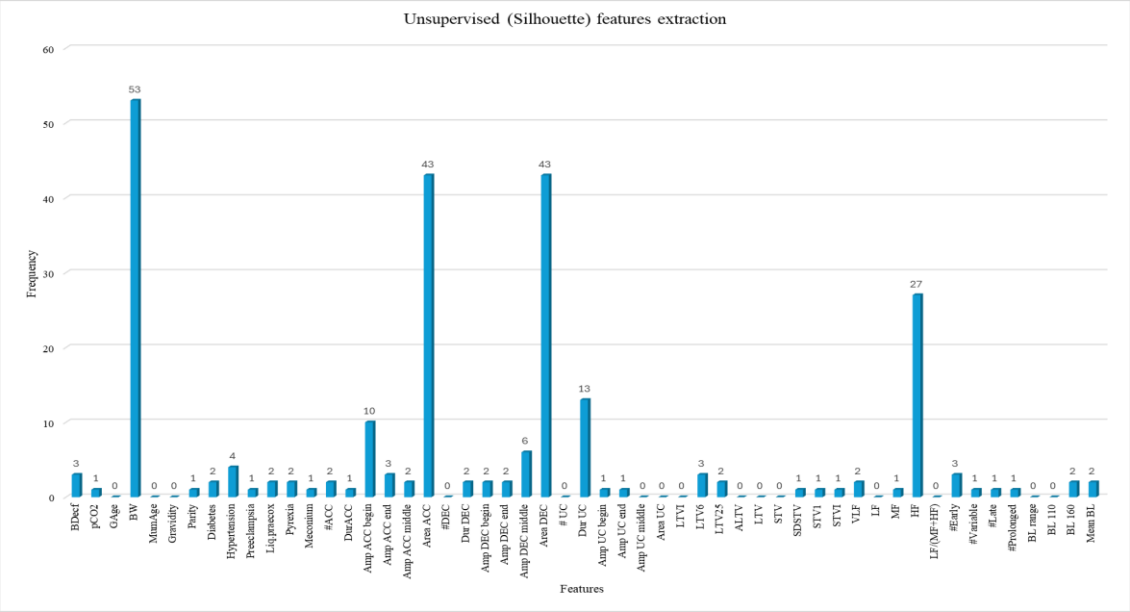

Figure 18 – Features extraction frequency via CTG-GCA model using the silhouette index as an evaluation metric

.

Figures 19 and 20 show the dendrograms in which the two clusters are represented with different colours (red and blue). In particular, Figure 19 represents the dendrogram formed by considering all the features of the database while Figure 20 represents the one constructed by taking into account the features extracted through the unsupervised CTG-GCA model, in which the silhouette index is used as an evaluation index of the clusters. At the end the clusters created considering all the features of the CTU-UHB database and in a second moment considering the extracted features were statistically evaluated and the results are reported in Tables 9.1, 9.2, 10.1 and 10.2. Each table shows the p-values for both clusters and the means and standard deviations for each feature of the database of both clusters. The values of the p values and the means and standard deviations of the features which appear to be significantly important for distinguishing the two groups were also highlighted.

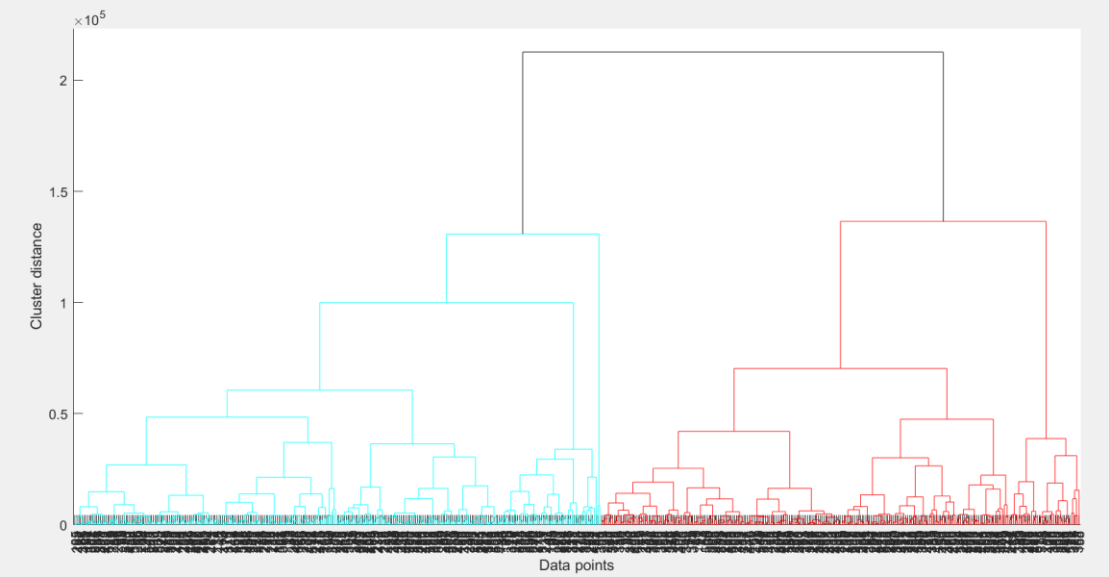

Figure 19 – Dendrogram and identificazione of the 2 clusters considering all the featured of the CTU-UHB database

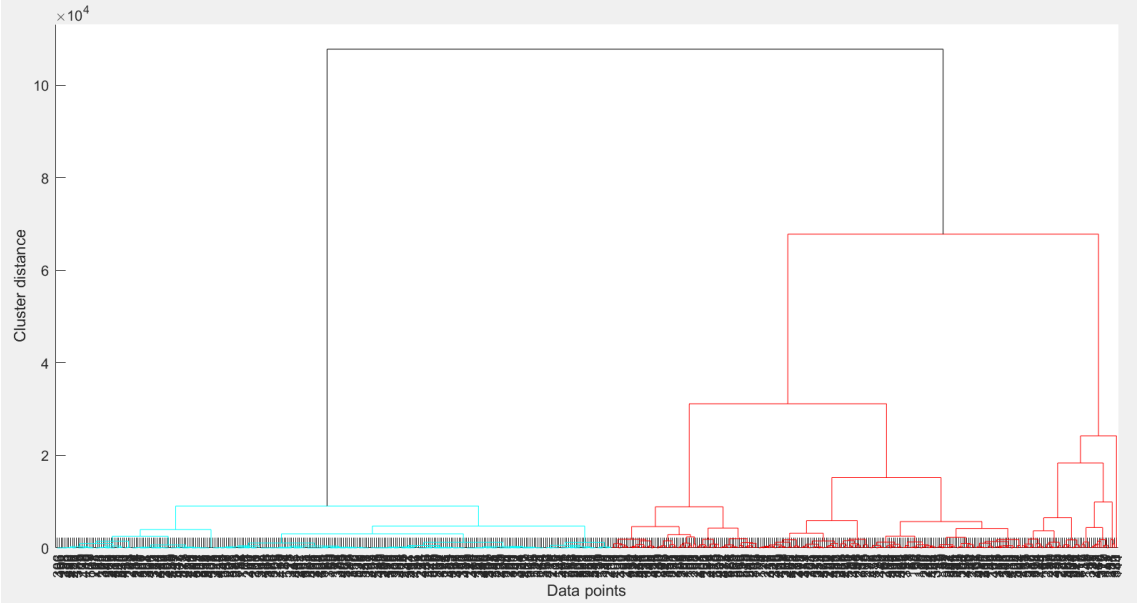

Figure 20 – Dendrogram and identification of the 2 clusters considering the features of the CTU-UHB database extracted with the unsupervised CTG-GCA model and with the Silhouette index as evaluation metric.

Features	Unit of measure	P-value		Mean ± Standard Deviation	
		Cluster 1	Cluster 2	Cluster 1	Cluster 2
pH	pH unit	0,2075	0,2075	7,224±0,111	7,237±0,098
Bdecf	mmol/L	0,1407	0,1407	4,789±3,935	4,204±3,113
pCO2	mmHg	0,6173	0,6173	6,918±1,927	6,94±1,723
BE	mmol/L	0,0787	0,0787	-6,583±4,244	-5,898±3,438
Apgar1	n.a.	0,8356	0,8356	8,273±1,624	8,251±1,629
Apgar2	n.a.	0,1993	0,1993	9,121±1,052	9,011±1,121
Gage	Weeks	0,4273	0,4273	40,021±1,112	39,966±1,157
BW	Years	0,7469	0,7469	3404,619±430,753	3382,89±523,471
Mum Age	n.a.	0,1592	0,1592	29,962±4,532	29,357±4,534
Gravidity	n.a.	0,2874	0,2874	1,422±0,99	1,369±1,083
Parity	n.a.	0,0836	0,0836	0,467±0,812	0,346±0,61
Diabetes	n.a.	0,8310	0,8310	0,069±0,254	0,065±0,246
Hypertension	n.a.	0,2130	0,2130	0,093±0,292	0,065±0,246
Preeclampsia	n.a.	0,3495	0,3495	0,024±0,154	0,038±0,192
Liq. Praecox	n.a.	0,7058	0,7058	0,26±0,439	0,274±0,447
Pyrexia	n.a.	**0,0099**	**0,0099**	**0±0**	**0,023±0,15**
Meconium	n.a.	0,2308	0,2308	0,1±0,301	0,133±0,34
ACC	n.a.	0,5884	0,5884	1,567±2,433	1,418±2,137
Dur ACC	seconds (s)	0,3725	0,3725	39,318±49,465	33,094±39,785
Amp ACC begin	bpm	0,1038	0,1038	16,351±18,68	13,78±15,766
Amp ACC end	bpm	**0,0298**	**0,0298**	**16,423±18,784**	**12,809±14,896**
Amp ACC middle	bpm	**0,0371**	**0,0371**	**16,176±18,531**	**12,908±15,108**
Area ACC	bpm*s	0,0713	0,0713	2583,97±3420,919	1754,912±2157,941
DEC	n.a.	0,2852	0,2852	4,055±4,889	4,133±4,423
Dur DEC	s	0,2276	0,2276	51,747±53,071	56,914±54,617
Amp DEC begin	bpm	**0,0053**	**0,0053**	**30,471±23,883**	**35,726±23,622**
Amp DEC end	bpm	**0,0115**	**0,0115**	**29,456±22,611**	**34,521±22,99**
Amp DEC middle	bpm	**0,0177**	**0,0177**	**29,54±22,679**	**34,082±22,711**
Area DEC	bpm*s	0,0713	0,0713	2583,97±3420,919	1754,912±2157,941
UC	n.a.	0,4224	0,4224	12,381±8,836	13,167±9,936
Dur UC	s	**0,0074**	**0,0074**	**62,15±29,996**	**57,349±30,98**
Amp UC begin	mmHg	0,0539	0,0539	42,66±23,2	38,814±21,823
Amp UC end	mmHg	**0,0304**	**0,0304**	**41,795±22,636**	**37,736±21,368**
Amp UC middle	mmHg	**0,0489**	**0,0489**	**41,984±22,717**	**38,225±21,562**
Area UC	mmHg*s	**0,0184**	**0,0184**	**5057,983±3511,245**	**4346,122±3007,184**
LTVI	ms	**1,80E-25**	**1,80E-25**	**453,661±5442,309**	**0,078±0,05**
LTV6	ms	**1,83E-45**	**1,83E-45**	**0,414±0,145**	**0,618±0,132**
LTV25	ms	**3,80E-05**	**3,80E-05**	**0,013±0,019**	**0,008±0,013**
ALTV	ms	**9,19E-84**	**9,19E-84**	**8736,82±3818,719**	**21468,79±7480,793**
LTV	ms	**1,26E-38**	**1,26E-38**	**8,132±1,709**	**6,124±1,444**
STV	ms	**1,84E-27**	**1,84E-27**	**8,242±2,807**	**6,007±2,066**

Table 9.1 – Results of statistical analysis of the 2 clusters created considering all the features. [1/2]

Features	Unit of measure	P-value		Mean ± Standard Deviation	
		Cluster 1	Cluster 2	Cluster 1	Cluster 2
SDSTV	ms	**7,91E-10**	**7,91E-10**	**9,808±3,254**	**8,2±3,102**
STV1	ms	**2,08E-45**	**2,08E-45**	**0,128±0,038**	**0,192±0,054**
STVI	ms	**3,92E-03**	**3,92E-03**	**0±0,001**	**0±0,001**
VLF	Hz	**9,99E-11**	**9,99E-11**	**0,032±0,037**	**0,018±0,02**
LF	Hz	**3,40E-17**	**3,40E-17**	**0,003±0,002**	**0,002±0,002**
MF	Hz	**2,63E-13**	**2,63E-13**	**0±0,001**	**0±0**
HF	Hz	**1,28E-10**	**1,28E-10**	**0±0,001**	**0±0**
LF(MF+HF)	n.a.	0,5540	0,5540	19,812±8,483	19,584±7,865
Early	n.a.	0,3973	0,3973	1,941±3,088	1,688±2,602
Variable	n.a.	**0,0079**	**0,0079**	**1,661±2,803**	**2,084±2,949**
Late	n.a.	0,7874	0,7874	0,353±0,946	0,304±0,776
Prolonged	n.a.	0,0976	0,0976	0,104±0,403	0,049±0,234
BL range	bpm	**9,68E-12**	**9,68E-12**	**14,557±7,828**	**10,946±6,691**
BL 110	s	**8,73E-07**	**8,73E-07**	**0,092±0,12**	**0,053±0,086**
BL 160	s	0,0687	0,0687	0,06±0,14	0,1±0,206
Mean BL	bpm	**9,02E-09**	**9,02E-09**	**133,207±11,703**	**139,047±12,139**

Table 9.2 - Results of statistical analysis of the 2 clusters created considering all the features. [2/2]

Features	Unit of measure	P-value		Mean ± Standard Deviation	
		Cluster 1	Cluster 2	Cluster 1	Cluster 2
pH	pH unit	0,5784	0,5784	7,231±0,111	7,229±0,099
Bdecf	mmol/L	0,0954	0,0954	4,47±3,907	4,555±3,18
pCO2	mmHg	0,6787	0,6787	6,869±1,948	6,995±1,695
BE	mmol/L	0,0816	0,0816	-6,195±4,239	-6,325±3,48
Apgar1	n.a.	0,2182	0,2182	8,294±1,698	8,228±1,544
Apgar2	n.a.	**0,0445**	**0,0445**	**9,121±1,141**	**9,011±1,021**
Gage	Weeks	0,1507	0,1507	40,048±1,12	39,935±1,146
BW	Years	0,7789	0,7789	3391,073±503,426	3397,776±446,788
Mum Age	n.a.	0,9255	0,9255	29,692±4,511	29,654±4,578
Gravidity	n.a.	0,8609	0,8609	1,429±1,122	1,361±0,93
Parity	n.a.	0,1375	0,1375	0,419±0,636	0,399±0,813
Diabetes	n.a.	0,6410	0,6410	0,062±0,242	0,072±0,259
Hypertension	n.a.	0,5225	0,5225	0,073±0,26	0,087±0,283
Preeclampsia	n.a.	0,9615	0,9615	0,031±0,174	0,03±0,172
Liq. Praecox	n.a.	0,1252	0,1252	0,239±0,427	0,297±0,458
Pyrexia	n.a.	0,4817	0,4817	0,014±0,117	0,008±0,087
Meconium	n.a.	0,5052	0,5052	0,107±0,31	0,125±0,332

Table 10.1 – Results of statistical analysis of the 2 clusters created taking into account the features selected with the unsupervised CTG- GCA model. [1/2]

Features	Unit of measure	P-value		Mean ± Standard Deviation	
		Cluster 1	Cluster 2	Cluster 1	Cluster 2
ACC	n.a.	**8,1651E-106**	**8,1651E-106**	**0,014±0,144**	**3,125±2,444**
Dur ACC	seconds (s)	**1,0351E-104**	**1,0351E-104**	**0,604±7,381**	**75,635±35,697**
Amp ACC begin	bpm	**8,6312E-106**	**8,6312E-106**	**0,128±1,56**	**31,607±10,589**
Amp ACC end	bpm	**2,6986E-105**	**2,6986E-105**	**0,111±1,838**	**30,733±10,957**
Amp ACC middle	bpm	**4,1806E-105**	**4,1806E-105**	**0,036±0,569**	**30,644±10,891**
Area ACC	bpm*s	**3,6414E-106**	**3,6414E-106**	**1,261±15,747**	**4592,947±2607,918**
DEC	n.a.	**7,8306E-08**	**7,8306E-08**	**3,218±4,1**	**5,053±5,059**
Dur DEC	s	0,0853	0,0853	53,77±64,457	54,691±39,06
Amp DEC begin	bpm	**0,0084**	**0,0084**	**29,961±26,03**	**36,285±20,826**
Amp DEC end	bpm	**0,0030**	**0,0030**	**28,633±24,664**	**35,425±20,274**
Amp DEC middle	bpm	**0,0058**	**0,0058**	**28,593±24,517**	**35,122±20,221**
Area DEC	bpm*s	**3,6414E-106**	**3,6414E-106**	**1,261±15,747**	**4592,947±2607,918**
UC	n.a.	0,1650	0,1650	12,266±9,356	13,293±9,385
Dur UC	s	0,4969	0,4969	60,042±32,025	59,664±28,87
Amp UC begin	mmHg	0,3284	0,3284	39,872±22,924	41,877±22,269
Amp UC end	mmHg	0,2086	0,2086	38,859±22,753	40,962±21,38
Amp UC middle	mmHg	0,2131	0,2131	39,097±22,519	41,397±21,896
Area UC	mmHg*s	0,5404	0,5404	4648,107±3318,373	4796,519±3278,088
LTVI	ms	0,5770	0,5770	226,874±3854,994	249,285±4041,055
LTV6	ms	**5,81655E-09**	**5,81655E-09**	**0,551±0,174**	**0,468±0,159**
LTV25	ms	0,3739	0,3739	0,011±0,019	0,01±0,015
ALTV	ms	0,6401	0,6401	15058,408±8961,275	14522,254±8292,191
LTV	ms	**2,01909E-08**	**2,01909E-08**	**6,797±1,948**	**7,591±1,705**
STV	ms	**0,0004**	**0,0004**	**6,897±2,787**	**7,486±2,615**
SDSTV	ms	0,5414	0,5414	8,979±3,388	9,111±3,162
STV1	ms	**2,24225E-05**	**2,24225E-05**	**0,168±0,06**	**0,148±0,05**
STVI	ms	0,7923	0,7923	0±0,001	0±0,001
VLF	Hz	0,3138	0,3138	0,024±0,027	0,027±0,035
LF	Hz	0,4471	0,4471	0,003±0,002	0,003±0,002
MF	Hz	0,2529	0,2529	0±0	0±0,001
HF	Hz	0,3991	0,3991	0±0	0±0,001
LF(MF+HF)	n.a.	0,2920	0,2920	20,081±8,418	19,289±7,922
Early	n.a.	**0,0027**	**0,0027**	**1,571±2,733**	**2,095±2,989**
Variable	n.a.	**2,40844E-10**	**2,40844E-10**	**1,246±2,411**	**2,54±3,187**
Late	n.a.	**0,0236**	**0,0236**	**0,294±0,924**	**0,369±0,804**
Prolonged	n.a.	0,0976	0,0976	0,104±0,403	0,049±0,234
BL range	bpm	**0,0376**	**0,0376**	**12,566±7,983**	**13,133±6,982**
BL 110	s	0,9641	0,9641	0,07±0,104	0,077±0,11
BL 160	s	**0,0319**	**0,0319**	**0,082±0,183**	**0,075±0,167**
Mean BL	bpm	0,1290	0,1290	136,71±12,143	135,198±12,351

Table 10.2 – Results of statistical analysis of the 2 clusters created taking into account the features selected with the unsupervised CTG-GCA model. [2/2]

3.3. Discussion

In this study is proposed the CTG-GCA model, an unsupervised ML algorithm, to extract the features from CTG signals taken from the CTU-UHB database. The selected features were then considered for the creation and formation of two clusters, one relating to hypoxic fetus and the other relating to the non-hypoxic ones.

The conception of the idea of this thesis and the creation of the unsupervised model arises from the fact that in literature, until now there has always been talks of classifying subjects into diseased or healthy using supervised ML techniques and it has never been mentioned the use of clustering techniques.

Compared to the supervised implementation of the proposed model, in which pH is used as gold standard, CTG-GCA promises good results.

From the histogram in Figure 18 it is possible to see how the selection of features is effective. In fact, out of 53 features, the model managed to extract 4, those with an extraction frequency greater than 20 %. The features extracted are BW (53%), Area ACC (43 %), Area DEC (43 %) and HF (27 %).

First positive feedback on the implementation of the model presented in the study can be obtained from the structure of the dendrograms in Figure 20. Compared to the dendrogram constructed considering all the features of the database (Figure 19), the one formed taking into consideration only the 4 extracted features (Figure 20) shows a clear distinction and distribution of the two clusters. Cluster 1 made up of 289 subjects and cluster 2 of 263.

The performances of the proposed model have not been compared to those models and studies already present in literature. Since this is an unsupervised ML algorithm it was not possible to apply the classic evaluation methods of supervised algorithms (accuracy, precision, recall, F1 score and AUC-ROC curve) based on the predictive accuracy with respect to the already known labels, but the statistical analysis was carried out.

In fact, from the results of the analysis, from the Tables 9.1, 9.2, 10.1 and 10.2, where the p-values less than 0.05 and the related mean and standard deviations, expressed in mean ± std dev, of the most significant features of each cluster have been highlighted, it is possible to find second positive feedback on the implementation of the proposed model.

From Tables 9.1 and 9.2, 27 features out of a total of 57 (47,36 %) were found to be significantly important in distinguishing between the two clusters. Among these it is possible to find features relating to accelerations (2), decelerations (4), UC (4), heart rate variability (13), baseline (3) and 1 feature relating to the clinical maternal information.
Similarly in Tables 10.1 and 10.2, 21 features out of 57 are shown to be significant, with a percentage of 36,84 %. Among these there are features relating to accelerations (6), decelerations (8), hear rate variability (4), baseline (2) and 1 referred to the clinical neonatal outcomes, Apgar 2.
The results are therefore very promising as they are similar to those of the supervised model with the addition of the features relating to neonatal clinical outcomes. This suggests that the 10-minute Apgar score is considered by the unsupervised model to construct the two clusters.
The Apgar index, that quickly assesses the child's physical condition and determines whether additional or emergency medical cares are needed immediately, is 9,121±1,141 in Cluster 1 and 9,011±1,021 in Cluster 2.
For this reason, this work, given the promising results obtained, can be considered a forerunner for this branch of research and a good start of point for future developments.
First of all, we could try to replace the clustering technique, using partition algorithms instead of hierarchical ones (dendrogram) as has been done up to now. A further variation that could be made could concern the evaluation metric within the fitness calculation; instead of evaluating the silhouette index, multivariate analysis of variance (MANOVA) could be implemented and verified.

Conclusion

This thesis proposes an analysis regarding the study of fetal characteristics in presence of fetal hypoxia. The detection, the diagnosis and an immediate intervention are vital for the health of the fetus and for this reason AI is increasingly called upon to provide support.
Due to the lack in literature of the use of unsupervised ML techniques for the evaluation of the fetus, in this thesis was proposed the implementation of the CTG-GCA.
The model presented is an unsupervised ML algorithm capable of selecting the relevant features in the identification of hypoxic fetus from non-hypoxic ones and creating and identifying the 2 clusters only by taking into consideration the features selected in the first step. To do this the genetic algorithm for the features selection and the hierarchical agglomerative clustering technique were used.
Using the genetic algorithm 4 features were extracted BW, Area ACC, Area DEC and HF and through the identification of the clusters, 2 were created. Cluster 1 composed of 289 subjects and Cluster 2 of 263.
From the statistical analysis carried out on them it emerged that the relevant characteristics in the identification of hypoxic and non-hypoxic fetus are not only the characteristics linked specifically to the CTG signal, such as accelerations, decelerations, uterine contractions, baseline or heart rate variability, but the Apgar index 5 minutes after birth is also significantly important.
This suggests that the Apgar index, as a neonatal clinical outcome, is able to define whether a fetus is hypoxic or not and this promising result can be seen and considered as a starting point for further future developments.
The study can therefore be considered innovative in the context of the use of unsupervised ML techniques for the evaluation of the fetus state assessment by cardiotocography.

Acknowledgments

I would like to express my deep gratitude and thanks to all the people who supported me in my university growth journey.
First of all, I would like to thank Dr. Agnese Sbrollini for her constant support, presence and patience in guiding me into the development of this study. His experience and enthusiasm have made this research journey an extraordinary education experience.
At the same time, I would like to thank Prof. Laura Burattini for giving me the opportunity to work in the Cardiovascular Bioengineering Lab. of the University.
A heartfelt thanks goes to my parents, Carlo and Sonia, to my brother Davide and to my family for their constant support and encouragement. For always celebrating my successes, for encouraging me not to give up and for pushing me to demonstrate how much I really was worth.
A special thank you goes to Luca, for always being on my side, for believing in me more than I did and for always pushing and encouraging me to give my best.
I would also like to thank my old friends and the new ones I met during my studies, thanks to whom I experienced moments of carefreeness that were essential to fully enjoy my university memories.
Thanks again to everyone who made this achievement possible.

Ancona, July 2024

Sara Capuano

Bibliography

[1] Ministero della Salute (http://www.salute.gov.it/portale/home.html)

[2] Glauco A. Dario C., Anatomia dell'uomo, Seconda edizione, Edi-Ermes, 2008

[3] Wood, C.E., and Keller-Wood, M.: 'Current paradigms and new perspectives on fetal hypoxia: implications for fetal brain development in late gestation', American Journal of Physiology-Regulatory, Integrative and Comparative Physiology, 2019, 317, (1), pp. R1-R13

[4] Thompson, L.P., Crimmins, S., Telugu, B.P., and Turan, S.: 'Intrauterine hypoxia: clinical consequences and therapeutic perspectives', Research and reports in neonatology, 2015, 5, pp. 79-89

[5] Omo-Aghoja L.: "Maternal and fetal Acid-base chemistry: a major determinant of perinatal outcome". Ann Med Health Sci Res. 2014 Jan;4(1):8-17

[6] Fanelli, Andrea, et al. "Quantitative Assessment of Fetal Well-Being through CTG Recordings: A New Parameter Based on Phase-Rectified Signal Average." IEEE Journal of Biomedical and Health Informatics, vol. 17, no. 5, Sept. 2013, pp. 959–66. PubMed.

[7] Sahin, Hakan. "Classification of the Cardiotocogram Data for Anticipation of Fetal Risks Using Machine Learning Techniques." APPLIED SOFT COMPUTING, vol. 33, Aug. 2015, pp. 231–38.

[8] Carbonne B., Benachi A., Lévèque ML, Cabrol D., Papiernik E., Maternal position during labor: effects on fetal oxygen saturation measured by pulse oximetry, Obstet Gynecol., 1996.

[9] Diogo Ayres-de-Campos, Catherine Y. Spong, Edwin Chandraharan, FIGO consensus guidelines on intrapartum fetal monitoring: Cardiotocography, International Journal of Gynecology and Obstetrics, 2015.

[10] Stephen E. Gee,Heather A. Frey, Contractions: Traditional concepts and their role in modern obstetrics , Seminars in Perinatology : 2020, 44(2), 151218

[11] Zafer Cömert, Abdulkadir Şengür, Yaman Akbulut, Ümit Budak, Adnan Fatih Kocamaz, Varun Bajaj, Efficient approach for digitization of the cardiotocography signals: 2020, 537(), 122725

[12] Ajirak M, Heiselman C, Quirk JG, Djurić PM. BOOST ENSEMBLE LEARNING FOR CLASSIFICATION OF CTG SIGNALS. Proc IEEE Int Conf Acoust Speech Signal Process. 2022 May;2022:1316-1320. Epub 2022 Apr 27.

[13] Santo S, Ayres-de-Campos D, Costa-Santos C, Schnettler W, Ugwumadu A, Da Graça LM, for the FM-Compare Collaboration. Agreement and accuracy using the FIGO, ACOG and NICE cardiotocography interpretation guidelines. Acta Obstet Gynecol Scand 2017; 96: 166–175.

[14] A. Sbrollini, S. Romagnoli, I. Marcantoni, L. Burattini, M. Morettini and L. Burattini, "Neonatal Clinical Outcomes: a Comparative Analysis," 2022 IEEE International Symposium on Medical Measurements and Applications (MeMeA), Messina, Italy, 2022, pp. 1-6

[15] M.J. Page, J.E. McKenzie, P.M. Bossuyt, I. Boutron, T.C. Hoffmann, C.D. Mulrow, L. Shamseer, J.M. Tetzlaff, E.A. Akl, and S.E. Brennan. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. 2020.

[16] Daydulo, Y.D., Thamineni, B.L., Dasari, H.K. et al. “Deep learning based fetal distress detection from time frequency representation of cardiotocogram signal using Morse wavelet: research study”. BMC Med Inform Decis Mak 22, 329 (2022)

[17] Ogasawara J, Ikenoue S, Yamamoto H, Sato M, Kasuga Y, Mitsukura Y, Ikegaya Y, Yasui M, Tanaka M, Ochiai D. “Deep neural network-based classification of cardiotocograms outperformed conventional algorithms.” Sci Rep. 2021 Jun 28;11(1):13367.

[18] F. Francis, H. Wu, S. Luz, R. Townsend and S. Stock, "Detecting Intrapartum Fetal Hypoxia from Cardiotocography Using Machine Learning," 2022 Computing in Cardiology (CinC), Tampere, Finland, 2022, pp. 1-4.

[19] Francis F, Luz S, Wu H, Townsend R, Stock SS. “Machine Learning to Classify Cardiotocography for Fetal Hypoxia Detection.” Annu Int Conf IEEE Eng Med Biol Soc. 2023 Jul;2023:1-4.

[20] V. Chud´aˇcek, J. Spilka, M. Burˇsa, P. Janku, L. Hruban, M. Huptych, and L. Lhotsk´a. Open access intrapartum ctg database. BMC pregnancy and childbirth, 14(1):1–12, 2014.

[21] A.L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C. K. Peng, and H. E. Stanley, “PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals.” *Circulation,* vol. 101, no. 23. Pp. E215-220, 2000.

[22] Sivanandam, SN e Deepa, SN. Introduction to genetic algorithms. s.l.: Springer,2007.

[23] Maad M.Mijwel, “Genetic Algorithm Optimazion by Natural Selection”, August, 2016

[24] Zhang Yang, Zhao Zhidong, Ye Haihui. “Valutazione intelligente dello stato fetale basata sull'algoritmo genetico e sulla macchina vettoriale di supporto dei minimi quadrati”, Journal of Biomedical Engineering, 2019, 36(1): 131-139.

[25] DK Rathore e PK Mannepalli, "A Review of Machine Learning Techniques and Applications for Health Care", Conferenza internazionale sui progressi nella

tecnologia, nella gestione e nell'istruzione (ICATME) del 2021, Bhopal, India, 2021, pp. 4-8.

[26] Janiesch, C., Zschech, P. & Heinrich, K. Apprendimento automatico e apprendimento profondo. Mercati degli elettroni 31, 685–695 (2021).

[27] Rui Xu and D. Wunsch, "Survey of clustering algorithms," in IEEE Transactions on Neural Networks, vol. 16, no. 3, pp. 645-678, May 2005.

[28] E. B., Cluster analysis. Chichester, West Sussex, U.K: Wiley, 2011.

[29] Peter J. Rousseeuw, “Silhouettes: A graphical aid to the interpretation and validation of cluster analysis”, Journal of Computational and Applied Mathematics, Volume 20, 1987, Pages 53-65.

[30] Warne, Russell, "A Primer on Multivariate Analysis of Variance (MANOVA) for Behavioral Scientists," Practical Assessment, Research, and Evaluation: Vol. 19, Article 17, 2019.