






## Article

# Machine Learning to Predict Pre-Eclampsia and Intrauterine Growth Restriction in Pregnant Women

Lola Gómez-Jemes <sup>1,†</sup> , Andreea Madalina Oprescu <sup>1</sup> , Ángel Chimenea-Toscano <sup>2</sup> , Lutgardo García-Díaz <sup>2</sup>   
and María del Carmen Romero-Ternero <sup>1,\*</sup> 

<sup>1</sup> Department of Electronic Technology, Universidad de Sevilla, 41012 Sevilla, Spain

<sup>2</sup> Hospital Universitario Virgen del Rocío, Departamento de Cirugía, Universidad de Sevilla, 41009 Sevilla, Spain

\* Correspondence: mcromerot@us.es

† Current address: ETSI Informática, Avda. Reina Mercedes s/n, 41012 Sevilla, Spain.

**Abstract:** The use of artificial intelligence in healthcare in general and in obstetrics and gynecology in particular has great potential. Specifically, machine learning methods could help improve the health and well-being of pregnant women, closely monitoring their health parameters during pregnancy, or reducing maternal and perinatal morbidity and mortality with early detection of pathologies. In this work, we propose a machine learning model to predict risk events in pregnancy, in particular the prediction of pre-eclampsia and intrauterine growth restriction, using Doppler measures of the uterine artery, sFlt-1, and PlGF values. For this purpose, we used a public dataset from a study carried out by the University Medical Center of Ljubljana, in which data were collected from 95 pregnant women with pre-eclampsia and intrauterine growth restriction. We adopted a multi-label approach to accomplish the prediction task. Different classifiers were evaluated and compared. The performance of each model was tested in terms of accuracy, precision, recall, F1 score, Hamming loss, and AUC-ROC. On the basis of these parameters, a variation of the decision tree classifier was found to be the best performing model. Our model had a robust recall metric (0.89) and an AUC ROC metric (0.87), taking into account the size of the data and the unbalance of the class.

**Keywords:** machine learning; multilabel classification; pre-eclampsia; intrauterine growth restriction; pregnancy disorders



**Citation:** Gómez-Jemes, L.; Oprescu, A.M.; Chimenea-Toscano, Á.; García-Díaz, L.; Romero-Ternero, M.d.C. Machine Learning to Predict Pre-Eclampsia and Intrauterine Growth Restriction in Pregnant Women. *Electronics* **2022**, *11*, 3240. <https://doi.org/10.3390/electronics11193240>

Academic Editor: Ahmad Taher Azar

Received: 5 August 2022

Accepted: 4 October 2022

Published: 9 October 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

In recent years, artificial intelligence (AI) has been increasingly applied in the fields of health and medicine. AI has great potential to help improve healthcare throughout the world. Some highlighted applications, according to a survey published in [1], could be the detection of hidden patterns in large volumes of healthcare data, analysis to aid clinical practice, and support healthcare professionals by providing up-to-date and trustworthy scientific information that can help reduce diagnosis errors and improve patient care. Furthermore, AI could be useful in developing countries and rural areas, where healthcare assistance may be limited or unavailable.

In the area of obstetrics and gynecology, the use of AI has shown increasing interest in the scientific community. In [2], authors reviewed the current state of research on methodologies, techniques, algorithms, and frameworks used in AI applied to pregnancy health and well-being. This study shows that AI can be applied to many pregnancy-related conditions or complications such as gestational diabetes, hypertension disorders, pre-eclampsia, preterm birth, mental health, and, in general, maternal and fetal well-being. In particular, machine learning (ML) has a wide range of applications in this field, including monitoring maternal and fetal health status, detecting risk factors during pregnancy, early detection of changes in a pathology, and prediction of preterm. ML can be a powerful tool that could

be used to support women during pregnancy and to improve maternal and fetal health status and well-being [2]. In particular, in the prevention of maternal risk during pregnancy, a growing number of studies show that ML can help as a prediction and detection tool. There are multiple topics of interest, such as real-time monitoring systems to detect changes in mother and fetus health status [3], prediction of gestational diabetes [4], prediction of postpartum hemorrhage [5], prediction of preterm [6], prediction of hypertension disorders such as HELLP syndrome (Hemolytic anemia, Elevated Liver enzyme, Low Platelet count) [7], and detection of abnormal image pattern on fetal ultrasound, such as congenital central nervous system (CNS) malformations [8].

In this work, we developed an ML model to detect placental dysfunction disorders. In particular, the model can predict if a pregnant woman suffers from pre-eclampsia (PE), intrauterine growth restriction (IGR), both, or none of the conditions.

PE and IGR are conditions related to placental insufficiency. On the one hand, pre-eclampsia is a specific pregnancy disorder that affects 3–5% of pregnancies worldwide [9]. It is a hypertension disorder that presents after 20 weeks of gestation. PE can be classified into early PE (before 34 weeks of gestation) and late-onset PE (after 34 weeks of gestation) [10]. Early-onset PE is commonly associated with other maternal organ dysfunctions, such as renal insufficiency, liver involvement, neurological or hematological complications, uteroplacental dysfunction, or fetal growth restriction. In contrast, late-onset PE is generally associated with mild disease, with a low impact on maternal and/or fetal outcomes [11]. On the other hand, the American College of Obstetricians and Gynecologists defines intrauterine growth restriction as “a fetus that fails to reach his/her potential growth” [12]. Infants with IURG have many acute neonatal problems that include perinatal asphyxia, hypothermia, hypoglycemia and polycythemia, and other long-term complications such as behavioral problems, cerebral palsy, growth failure, and lower levels of intelligence, among others [13]. Both PE and IGR are considered important causes of maternal, neonatal, and fetal morbidity and mortality [14]. Being able to predict these conditions early in pregnancy would be crucial to improving newborn and maternal outcomes. Therefore, the development of a ML model to predict these diseases could be a very valuable tool to support clinicians in making decisions.

Pre-eclampsia and intrauterine growth restriction are characterized by abnormal placental formation that results in inadequate uteroplacental blood flow [15]. Uterine Artery Doppler ultrasound is a non-invasive diagnostic method that uses high-frequency sound to assess the uteroplacental circulation. The use of Doppler of the uterine artery has not been accepted in routine practice, but in combination with the angiogenic markers sFlt-1 (soluble fms-like tyrosine kinase-1) and PlGF (placental growth factor), it could become a very powerful tool for the prediction and early diagnosis of pre-eclampsia and intrauterine growth restriction [11].

In the scientific literature, previous work on this topic uses this dataset to predict PE and IGR [16]. The main shortfall of their model is that it only solves a classification task between a control group and a group with placental dysfunction-related disorder (PDD), pre-eclampsia or IUGR. Therefore, the authors do not differentiate whether a pregnant woman has only pre-eclampsia, IURG, or both conditions. This paper proposes a new approach: a multi-label classification. The principal characteristic of our model is that it can predict whether a pregnant woman suffers from PE, IURG, both disorders, or none of them.

## 2. Materials and Methods

### 2.1. Model Design

We have developed a machine learning model to predict pregnancy outcomes. To report its results, we have followed the guidelines specified in the Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research: A Multidisciplinary View [17]. Python was the programming language chosen to develop the machine learning model. Scikit-learn was used to implement the ML algorithm. Scikit-learn [18] is

a machine learning library written in Python. It provides a wide range of state-of-the-art machine learning algorithms for supervised (including the multi-output classification and regression algorithm) and unsupervised problems.

We used a public dataset from a prospective cohort study on the use of Doppler measures of the uterine arteries and the sFlt-1/PIGF ratio in hypertensive disorders during pregnancy [19]. The model was designed to make a prognosis of pregnancy outcomes. In particular, the model had to solve a multi-label classification task. The under-prediction of the model can increase maternal and neonatal mortality and morbidity, while the over-prediction can increase health care costs. We aim to avoid both scenarios, but prioritizing the prediction of pregnant women with a placental dysfunction disorder. The metrics used to evaluate the performance of the model were precision, recall, F1 score, AUR-ROC, Hamming loss, and confusion matrix. In addition, a Dummy Classifier was used as a baseline model. We defined a classifier that made predictions based on the most frequent label of the dataset. All code developed for the model is available in the Appendix A.

## 2.2. Dataset

The dataset used is publicly available in Mendeley Data [19]. These data belong to a study conducted by the University Medical Center Ljubljana, from September 2012 to January 2015 [20]. The study was approved by the Republic of Slovenia National Medical Ethics Committee (No. 104/04/12). Data were collected from 95 patients with a singleton pregnancy between 24 and 38 weeks of gestation. The study sample included 22 women with PE, 32 women with PE and IGR, 12 women with IGR, and 29 women with low-risk pregnancy as a control group (without any signs of hypertensive disorders during pregnancy, pre-pregnancy hypertension, pre-pregnancy diabetes, or gestational diabetes). The features provided in the dataset included maternal characteristics, neonatal characteristics, Doppler measures of the uterine arteries (for the right and left uterine artery), sFlt-1 value, PIGF and the ratio of sFlt-1/PIGF. Mean values were also included for each measure of the uterine artery. All features are listed and described in Table 1.

**Table 1.** Features available in the dataset [19].

Feature	Description
Class	Target variable. Patient health status at the time of data collection. Four possible classes: control (low-risk pregnancy), PE (only pre-eclampsia), IUGR (only early-onset uterine growth restriction), and IUGR + PE (both PE and IGR)
<b>Neonatal Characteristics</b>	
Weight	Neonatal weight in grams
<b>Maternal Characteristics</b>	
Maternal age	Patient age
Parity	Number of times that a woman has delivered a fetus with a gestational age of 24 weeks or more, regardless of whether the child was born alive or was stillborn
Pre-pregnancy weight	Weight of a woman before pregnancy, in kilos
Maternal Height	Patient height in meters
BMI before pregnancy	Body mass index before pregnancy. It is calculated by dividing the weight by the square height. Unit: kg/m <sup>2</sup>
Gestational age at delivery	Gestational age at delivery in weeks

Table 1. Cont.

Feature	Description
<b>S-Flt1 and PlGF Measures</b>	
S-Flt1	Serum levels of fms-like soluble tyrosine kinase. Unit: 1 µg/L
S-PlGF sFLT/PLGF	Placental growth factor µg/L sFlt-1 and PlGF ratio
<b>Uterine Arteries Doppler (UtAD) Measures</b>	
Art ut. D-resistance index [RI]	Resistance index of the right uterine artery
Art ut. L-resistance index [RI]	Resistance index of the left uterine artery
Mean RI	Average resistance index
Art ut. D-pulsatility index [PI]	Pulsatility index of the right uterine artery
Art ut. L-pulsatility index [PI]	Pulsatility index of the left uterine artery
Mean PI	Average Pulsatility Index
Art ut. D-Peak Systolic Velocity [PSV]	Peak systolic of the right uterine artery
Art ut. L-Peak Systolic Velocity [PSV]	Peak systolic of the left uterine artery
Mean PSV	Average peak systolic
Bilateral notch	Presence of notch. Three possible values: 2: both arteries have a notch, 1: an artery has a notch (right or left), 0: no notch detected

Initially, gestational age at delivery and weight were discarded. These features are collected at the end of pregnancy. Therefore, they cannot be used to detect risks during pregnancy.

### 2.3. Exploratory Analysis

The dataset contains 95 instances and 21 features. As mentioned above, we initially dismissed gestational age at delivery and weight. We also removed the ID column. Seven null values were found in the dataset: three values in the BMI column, two values in mean PSV, one value in pre-pregnancy weight, and one value in height. These values could be inferred in the preprocessing stage. The target variable was multiclass. There were four different categories: IUGR\_PE, Control, PE, and IUGR. The distribution of these features was imbalanced, as can be observed in Figure 1.

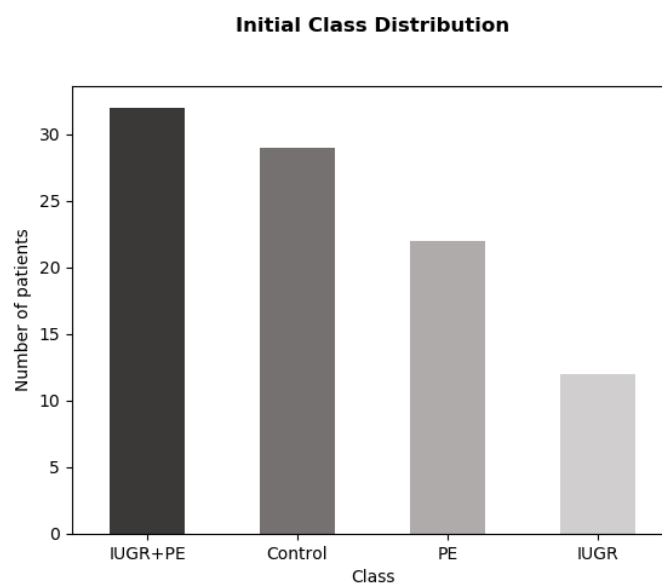


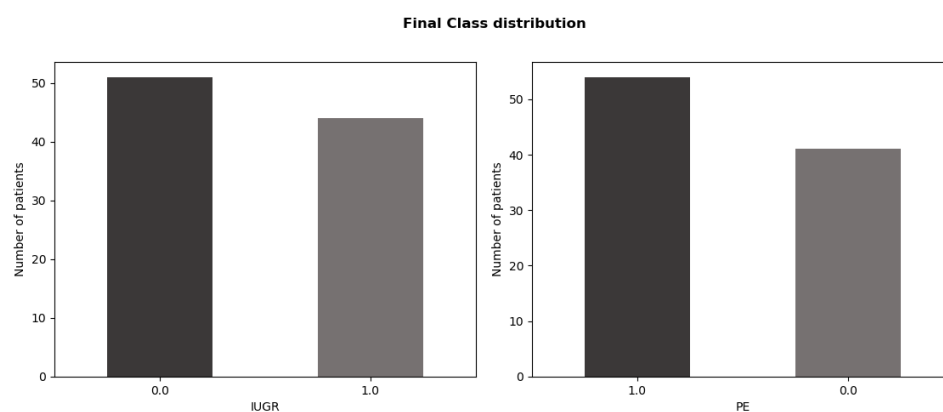
Figure 1. Initial distribution of the class.

During the exploratory data analysis, it was detected that the problem could be treated as a multi-label instead of a multi-class classification. The IUGR + PE and Control classes truly depended on the presence or absence of PE or IUGR. Thus, we define two binary tags as output: PE and IUGR. The encoding of each category is described in Table 2.

**Table 2.** Possible output of the multi-label classification model.

PE	IUGR	Meaning
0	0	Baseline
1	0	Pre-eclampsia
0	1	Intrauterine growth restriction
1	1	Both

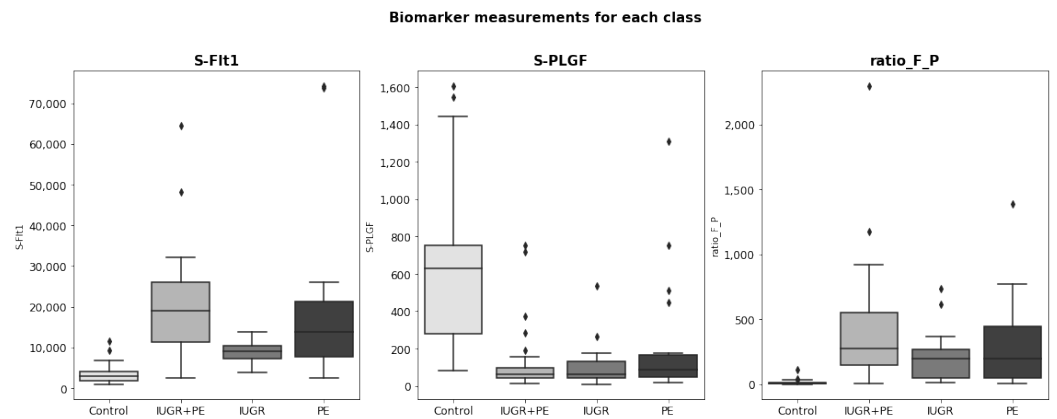
The new distribution of the target variable was more balanced, as shown in Figure 2.



**Figure 2.** Final distribution of the target variable. The class was transform into two binary label: PE and IURG. The distribution of each label is more balanced than the initial class distribution.

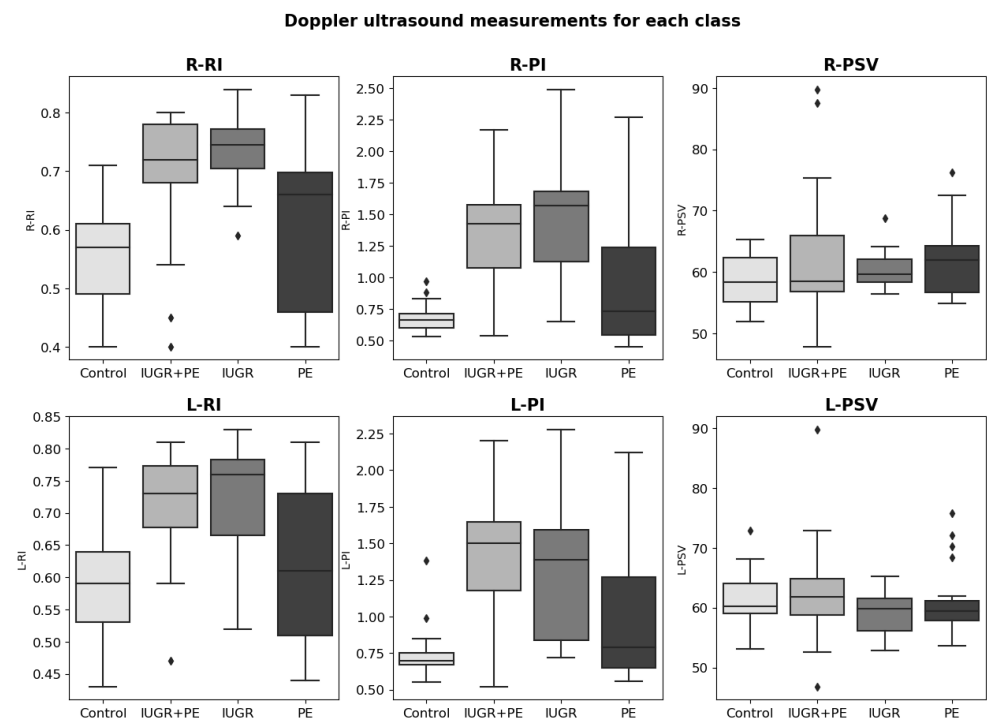
All characteristics were numerical, except parity and bilateral notch, which were considered categorical variables. The numerical variables had different magnitudes (for example, the PIGF value is three orders of magnitude higher than the height value). In addition, many features had a skewed distribution. Doppler ultrasound measurements and the biomarkers sFlt-1 and PIGF are relevant in the early detection of PE and IRG. We studied the distribution of these variables for each class to prove their relevance.

For all biomarker distributions, there were significant differences between the control group and the groups with any placental disorder (see Figure 3). Specifically, the sFlt-1 distribution of the control group is lower than the rest, while the PIGF value is much higher. Among the placental disorder groups, notable differences in the sFlt-1 value were also detected (in the case of PIGF, all shared a similar distribution).

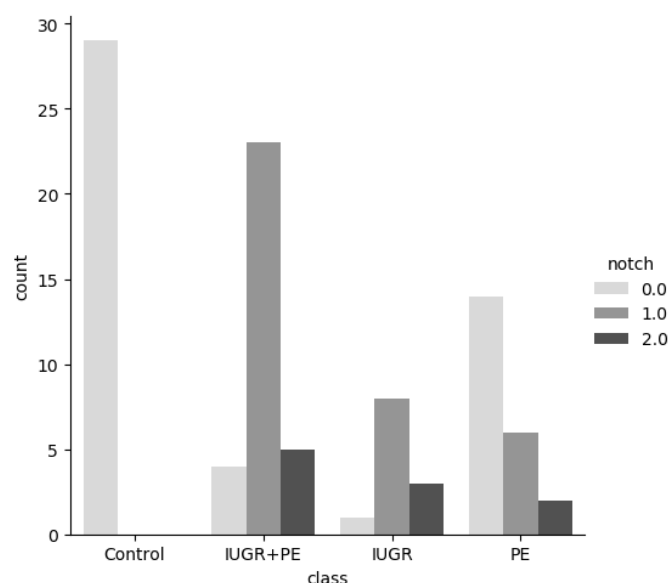


**Figure 3.** Distribution of sFlt-1, PIGF, and sFlt-1/PIGF ratio by class. The lozenge symbol ♦ represents the outliers values of each measure.

In Doppler ultrasound measurements (see Figure 4), we also observed important differences between categories in the Pulsatility Index value and the Resistance Index value. For the PSV value, the differences were less significant. Regarding the notch feature, it was observed that no notch was found in the arteries in the control group (see Figure 5). Furthermore, we could see that unilateral notch is present in a higher proportion in pregnant women with IGR.



**Figure 4.** Distribution of UtAD measures for left (L) and right (R) uterine artery by class. The lozenge symbol “♦” represents the outliers values of each measure.



**Figure 5.** Distribution of notch feature by class: 2.0 means that both arteries have a notch, 1.0 that an artery has a notch (right or left), and 0.0 that no notch has been detected.

Finally, we study the correlation between the variables. Several highly correlated variables were detected. The features meanPI, meanRI, meanPSV, and BMI contained information about other columns (for example, meanRI contains the mean value of the resistance index of the left and right). Therefore, they were very closely related to other columns.

#### 2.4. Data Preprocessing

Several transformations were applied to the raw data. First, all high-correlated features were dropped (meanRI, meanPI, meanPSV, and BMI). If two predictors are highly correlated, this implies that they are measuring the same underlying information. Removing one should not compromise the performance of the model and might lead to a more interpretable model. Even some models can improve their performance by removing these variables [21]. After removing said variables, only two null values remained, one in height and the other in weight column. We calculate the relative standard deviation (RSD) to see whether the average could be a representative value of the features. RSD is obtained by dividing the standard deviation by the average and dividing by 100 (it is expressed as a percent). The RSDs were 3.78% and 21.26% for height and weight, respectively. Thus, we concluded that the data were clustered around the mean and that we could use it to impute the missing values. Concerning the numerical features, two transformations were applied: logarithmic transformation and standardization. The logarithmic transformation was applied to features with a skewed distribution. Replacing the data with the log can help remove the skew [21]. Standardization was applied to all numerical variables to homogenize their magnitudes. Categorical features were encoded. We used the one-hot encoding technique, which consists of creating as many columns as different values are contained in the column and attributing the value 1 to the category to which the data correspond, and 0 to the rest of them. Finally, the target variable was also encoded. We defined a function to transform the target variable into two binary features: PE and IURG. Data were divided into test (20%) and training (80%) sets. The training set was used to train the models and the test set was used for validation.

#### 2.5. Model Training

Learning from multi-label data can be achieved through different approaches, such as data transformation, adaptation of traditional classification methods, and use of ensembles



of classifiers [22]. In this work, we focus on the data transformation and method adaptation approach.

The data transformation method is based on transformation techniques that transform the original multilabel data into one or more binary or multiclass datasets. On the other hand, the adaptation method consists of adapting existing classification algorithms, so that they can process multi-label data and produce several outputs instead of one [22]. Some models that can be adapted to multilabel classification are Decision Tree Classifier, Extra Tree Classifier, Random Forest Classifier, and K-Nearest Neighbors Classifier.

Decision tree is a non-parametric supervised learning algorithm which can be used to solve classification task. It has a tree structure consisting of a root node, branches, internal nodes, and leaf nodes. It employs a divide-and-conquer strategy, which is a recursive partitioning of the problem into two or more subproblems until it becomes simple enough to be solved directly. Thus, the decision tree classifier splits the data in a top-down, recursive manner until all, or the majority of records have been classified under the specific class labels. As parameters, we have selected the Gini impurity (the probability of misclassifying an observation) to measure the quality of a split, the best split at each node as a split criterion, two minimum number of samples to split an internal node, and at least one sample to be at a leaf node.

The Extra Tree Classifier is a variation of a Decision Tree Classifier. It consists of an extremely randomized tree classifier. It strongly randomizes both the attribute and the cut-point choice while splitting a tree node. As parameters, we have selected the Gini impurity (the probability of misclassifying an observation) to measure the quality of a split, the random split at each node as a split criterion, two minimum numbers of samples to split an internal node, and at least one sample to be at a leaf node.

Random Forest Classifier is also a tree-based method that consists of a large number of individual decision trees that operate as an ensemble. It is an extension of the bagging method as it utilizes both bagging and feature randomness to create an uncorrelated forest of decision trees. In the bagging method, a random sample of data is selected from a training set for replacement. Then, several data samples are generated and they are used to train the models independently. The feature randomness (also known as the random subspace method) generated a random subset of features, which ensures the low correlation between the different decision trees generated. This is an important difference between decision trees and random forests: while decision trees consider all the possible feature splits, random forests only select a subset of those features. In the classification task, the output of the random forest model is the most voted class among all decision trees. As hyperparameters, we chose two as node size, and the number of trees in the forest was set to one hundred. To set the number of features to use, we used the square root of the total of features.

Finally, the K-Nearest Neighbors (KNN) Classifier is an instance-based learning algorithm. It is a lazy learning algorithm, as it delays the induction or generalization process until classification is performed. KNN algorithm assumes that instances within a dataset will generally exist in close proximity to other instances that have similar properties. KNN works by finding the distances between an unclassified instance and all the instances in the data, selecting the specified number of examples (K) closest to it, and then determining its label by identifying the most frequent label of its neighbors. To calculate the distance between the instances, we used the Euclidean distance. Moreover, the K value chosen was five.

Regarding data transformation, we applied two of the methods proposed in the literature: Binary Relevance and Label Powerset. Binary relevance is a straightforward approach to handling a multilabel classification task. It decomposes the learning of each label into a set of binary classification tasks, one per label, where each model is independently learned, using only the information from that label and ignoring the information from the others [23]. The main drawback of this technique is that it does not consider any label dependency. However, this technique also has advantages, such as that any binary learning algorithm



can be used as an estimator, and it has linear complexity for the number of labels [23]. In our case, we used Gaussian Naïve Bayes, Random Forest Classifier, Support Vector Machine, K Neighbors Classifier, and Decision Tree Classifier as estimators.

The label powerset method proposes using each different combination of labels as an identifier of a new class. The resulting dataset has only one class. Thus, it can be treated as a multiclass classification [24]. We used as an estimator the Random Forest Classifier, Support Vector Machine, K Neighbors Classifier, and Decision Tree Classifier. We tested this method to evaluate the performance of models considering a multiclass problem (the original type of problem). A training set was used to train all these models.

### 2.6. Model Validation

After training the models, we use the test set to evaluate their performance. Model selection was evaluated by the AUC ROC (area under the curve ROC), accuracy (fraction of instances that the model classified correctly), precision (proportion of positive identifications that were actually positive), recall (proportion of the positive class that was correctly classified), F1 score (harmonic mean of precision and recall), and Hamming loss (proportion of misclassifications). Label-based measures decompose the evaluation of each label. There are two options available: averaging the measure label-wise (macro-average) or concatenating all label predictions and computing a single value over all of them (micro-average). Macro-average will compute the metric independently for each label, so it gives equal weight to all labels. On the contrary, micro-average metrics aggregate the contributions of all labels to compute the average metric [23]. We use the macro-average version of recall, precision, and F1 score.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

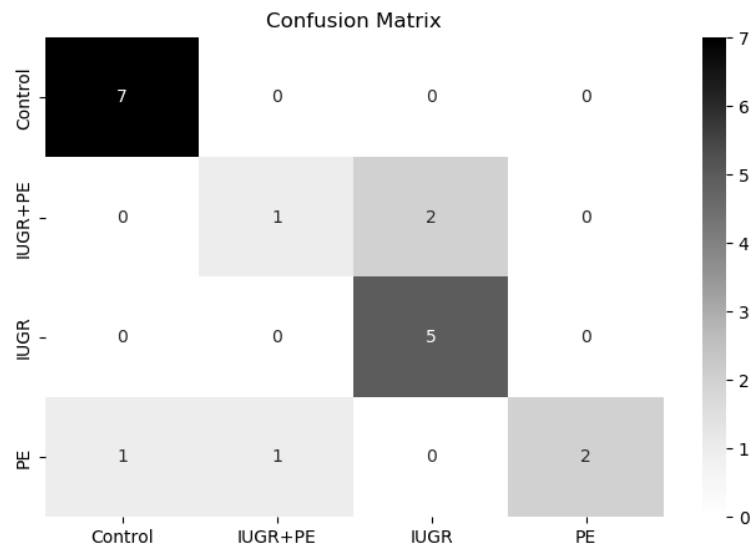
$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1\ score = \frac{2 * TP}{2 * TP + FP + FN} \quad (4)$$

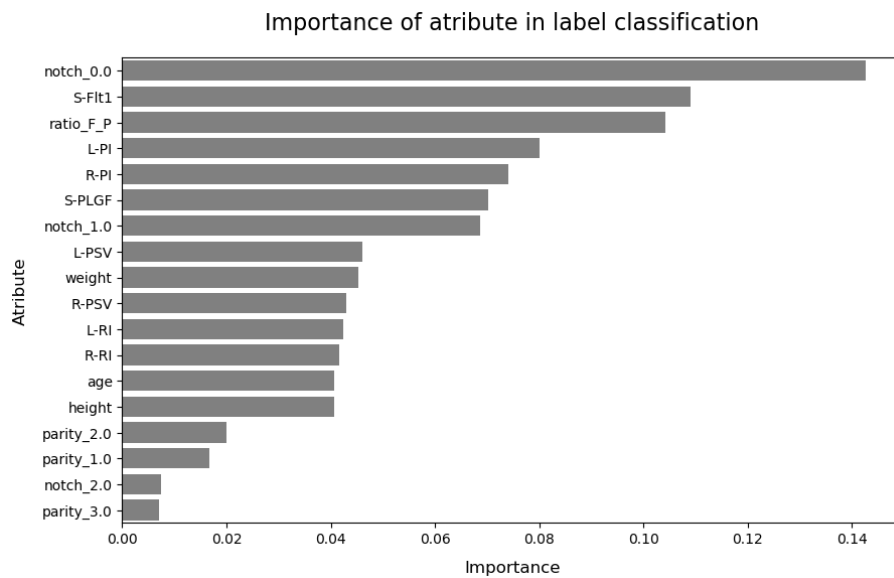
## 3. Results

The extra tree classifier model was found to achieve the best performance metrics of all models (see Table 3), as determined by the AUC ROC value. The higher the AUC ROC, the better the model is in discerning between patients with any placental disorder and without any disorder. The model achieved 0.789474 in accuracy, 0.83333 in precision, 0.888889 in recall, 0.859477 in the F1 score, 0.871717 in the AUC ROC, and 0.131579 in Hamming loss. These metrics were better than the baseline model metrics. We decoded the output of the model and calculated the confusion matrix to see how good the prediction of the classes was. As can be seen in Figure 6, the model sometimes failed to predict both disorders: it only predicted one of them. The model needs to be trained with more data from pregnant women with both disorders, so it could improve its performance.



**Figure 6.** Confusion matrix for the extra-tree classifier model. The multi-label output was decoded to see the prediction performance for each class.

We studied the importance of each feature in label classification (Figure 7). They were computed as the (normalized) total reduction in the criterion brought by that feature (Gini importance). The absence of a notch was found to be the most important characteristic, followed by the value of the S-Flt1 and sFlt-1/PlGF ratio. This result is consistent with what has been reviewed in the scientific literature: Doppler measures and the ratio sFlt-1/PlGF are important indicators for predicting pre-eclampsia and intrauterine growth restriction. Furthermore, it can also be observed that maternal characteristics were less relevant in the classification task.



**Figure 7.** Importance of the extra tree classifier characteristic.

**Table 3.** Performance metrics for all models evaluated, sorted by the AUC-ROC measure.

Model	Accuracy	Precision	Recall	F1 Score	AUC ROC	Hamming Loss
Extra Trees	0.789474	0.833333	0.888889	0.859477	0.871717	0.131579
Random Forest	0.736842	0.826389	0.826389	0.826389	0.840467	0.157895
Binary relevance—Random Forest	0.736842	0.826389	0.826389	0.826389	0.840467	0.157895
Label Powerset—SVC	0.631579	0.752137	0.944444	0.834225	0.824495	0.184211
Binary Relevance—Gaussian NB	0.631579	0.777778	0.833333	0.803922	0.818939	0.184211
Label Powerset—Random Forest	0.631579	0.850000	0.763889	0.796992	0.806944	0.184211
Binary Relevance—SVC	0.631579	0.718182	0.888889	0.794444	0.798990	0.210526
Binary Relevance—K Neighbors Classifier	0.578947	0.755682	0.826389	0.787500	0.790467	0.210526
K Neighbors	0.578947	0.755682	0.826389	0.787500	0.790467	0.210526
Label Powerset—K Neighbors Classifier	0.578947	0.778571	0.763889	0.768421	0.784217	0.210526
Binary Relevance—Decision Tree Classifier	0.526316	0.658654	0.812500	0.721591	0.738068	0.263158
Label Powerset—Decision Tree Classifier	0.526316	0.725000	0.576389	0.618421	0.690467	0.289474
Decision Tree	0.421053	0.651515	0.638889	0.635714	0.673990	0.315789
Dummy Classifier	0.210526	0.236842	0.500000	0.321429	0.500000	0.473684

#### 4. Discussion

We developed an ML model to predict pre-eclampsia and uterine growth restriction using data from pregnant women at 24–37 weeks of gestation. These data included maternal and fetal characteristics, as well as Doppler measures of the uterine artery, sFlt-1, and PlGF values.

Recently, the term placental dysfunction-related disorder (PDD) has been implemented to include two entities with a common etiopathogenic origin: pre-eclampsia and IUGR. Although IUGR is one of the leading causes of fetal morbidity and mortality, pre-eclampsia is associated with hypertension and multiorgan dysfunction, being one of the leading causes of death in pregnant women worldwide [25,26]. Its importance lies not only in its severity but also in its high prevalence, which can affect up to 5% pregnant women [27]. Every year, 500,000 babies and 76,000 women die in the world from these disorders.

Prediction of these entities can change the course of the disease, as these strategies will allow follow-up to anticipate and recognize the onset of the clinical syndrome and prevent or mitigate the development of PDD. Thus, in the scientific literature, models have generally been developed aimed at predicting PDD in the first trimester of pregnancy with maternal risk factors and biomarkers as a one-step procedure [28]. Several studies have shown that low-dose aspirin initiated at <16 weeks' gestation can be effective in reducing the prevalence of early-onset PE 8 with delivery at <34 + 0 week's gestation) [29–31], and also fetal growth restriction [30].

Although some research has reported optimal results only when treatment begins before 16 weeks, The American College of Obstetricians and Gynecologists and the Society of Maternal-Fetal Medicine support that low-dose aspirin should be started between 12 weeks and 28 weeks of gestation and continued daily until delivery [29].

Therefore, pregnant women who have not been deemed at high risk for PDD during the first trimester screening could benefit from this model: PDD could be predicted during second-trimester screening and prescribe aspirin or start monitoring the evolution of the condition. Prediction of placental disease at 20 weeks opens a window of opportunity for those pregnant women who have not been able to receive adequate counseling and treatment in the first trimester.

This second-trimester prediction is particularly important in low- and/or middle-income countries (LMIC), due to a number of barriers that limit first trimester care, delayed first antenatal visit, or even contact with a healthcare worker. Furthermore, maternal mortality from pre-eclampsia is highest in LMIC, and pregnant women are at a higher risk

of developing PDD [15]. About 99% of serious morbidity occurs in LMIC, which makes prediction and prevention especially important in these countries.

Many barriers and factors can contribute to the low adoption of an early antenatal care visit in LMIC. Lack of knowledge, socioeconomic status, availability, accessibility, acceptability, family support, and previous experiences with the health system affect the timing of the first visit [32]. In 2013, the estimated coverage of early antenatal care visits was 24% in low-income countries compared to 81.9% in high-income countries [32]. In this context, many pregnant women do not have the opportunity to access early pre-eclampsia screening during the first trimester of pregnancy. Therefore, they could benefit from a model trained with data from the second trimester of gestation.

Our model has limitations that are derived mainly from the limited number of samples in the dataset used. Using a small dataset to train and test a prediction model might lead to an overestimation of performance. Although the amount of health data that can be collected is rapidly increasing, the availability of large publicly available datasets is still limited to researchers. Sharing health data presents multiple challenges, including integration, ethics, privacy, and regulations, among others.

External validation is needed to confirm the predictive performance of the model. Additionally, more studies are required to determine whether the integration of other predictive clinical characteristics into the model could improve its performance and generalization. More efforts must be made to incorporate this application into clinical practice. There are a variety of technical challenges. In addition to the challenge of lack of data, there is currently a lack of interoperability standards in terms of data structures in the databases of each hospital and practice. Another challenge is building trust towards AI solutions for all stakeholders: patients, medical practitioners, and managers. To achieve a mutual benefit relationship, maintaining a human-centered design is key.

## 5. Conclusions

In this article, we have developed a machine learning model to predict risk during pregnancy, in particular pre-eclampsia and intrauterine growth restriction. Both are disorders of placental dysfunction that are an important cause of maternal, neonatal, and fetal morbidity and mortality. Their detection could help improve newborn and maternal outcomes.

The extra tree classifier has achieved the best metric of all the models evaluated in terms of AUC ROC (0.87). It has a robust performance in classifying different placental disorders versus a control group. However, the model sometimes fails to detect pregnant women with both disorders. However, we demonstrate that a simple classification model performs quite well, and we consider that it could be used as a baseline classifier model to continue improving the prediction of pre-eclampsia and intrauterine growth restriction.

Moving forward with research, we strongly encourage researchers to contribute to open health data, sharing, when possible, anonymised health data. Furthermore, if a programming code has been used to reach the results in the paper, we encourage researchers to share it, in order to improve research reproducibility. We also encourage researchers to develop and implement new interpretable machine learning methods for health research, as a means to contribute to fair and ethical decision-making, which can lead to building trust in AI.

In conclusion, this article shows how machine learning could be used to improve maternal and fetal health and well-being, as well as to support women during such a complex vital period.

**Author Contributions:** Conceptualization and methodology: L.G.-J., A.M.O. and M.d.C.R.-T.; data curation: L.G.-J.; formal analysis: L.G.-J., A.M.O., Á.C.-T., L.G.-D. and M.d.C.R.-T.; investigation: L.G.-J., A.M.O., Á.C.-T., L.G.-D. and M.d.C.R.-T.; software: L.G.-J.; writing—original draft preparation: L.G.-J.; writing—review and editing: A.M.O., Á.C.-T., L.G.-D. and M.d.C.R.-T.; supervision: M.d.C.R.-T. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are openly available in Mendeley Data at <https://doi.org/10.17632/zsjhvy9ytx.1>.

**Acknowledgments:** We would like to thank for the open dataset available on Mendeley Data to Tanja Premru-Srsen from the Department of Perinatology, Division of Obstetrics and Gynecology, University Medical Centre Ljubljana, Slovenia.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
AUC	Area under the receiver operating characteristic curve
BMI	Body mass index
CNS	Central nervous system
HELLP	Hemolytic anemia, Elevated Liver enzyme, Low Platelet count
IUGR/IGR	Intrauterine growth restriction
KNN	K-Nearest Neighbour
LMIC	Low and/or middle income countries
ML	Machine learning
PDD	Placental dysfunction-related disorder
PE	Pre-eclampsia
PI	Pulsatility Index
PIGF	Placental growth factor
PSV	Peak Systolic Velocity
RI	Resistance Index
ROC	Receiver operating characteristic
RSD	Relative standard deviation
sFlt-1	Soluble fms-like tyrosine kinase receptor-1
UtAD	Uterine Arteries Doppler

## Appendix A

All developed code is publicly available in GitHub (<https://github.com/lolagi/TFM-Modelo-Riesgos-Embarazadas> (accessed on 3 October 2022)). The academic research (master's thesis) that served as the base for this article is also available in the GitHub repository.

## References

- Jiang, F.; Jiang, Y.; Zhi, H.; Dong, Y.; Li, H.; Ma, S.; Wang, Y.; Dong, Q.; Shen, H.; Wang, Y. Artificial intelligence in healthcare: Past, present and future. *Stroke Vasc. Neurol.* **2017**, *2*, 230–243. [[CrossRef](#)] [[PubMed](#)]
- Oprescu, A.M.; Miró-Amarante, G.; García-Díaz, L.; Beltrán, L.M.; Rey, V.E.; Romero-Ternero, M. Artificial Intelligence in Pregnancy: A Scoping Review. *IEEE Access* **2020**, *8*, 181450–181484. [[CrossRef](#)]
- Veena, S.; Aravindhar, D.J. Remote Monitoring System for the Detection of Prenatal Risk in a Pregnant Woman. *Wirel. Pers. Commun.* **2021**, *119*, 1051–1064. [[CrossRef](#)]
- Hou, F.; Cheng, Z.; Kang, L.; Zheng, W. Prediction of Gestational Diabetes Based on LightGBM. In Proceedings of the 2020 Conference on Artificial Intelligence and Healthcare, Taiyuan, China, 23–25 October 2020; Association for Computing Machinery: New York, NY, USA, 2020; pp. 161–165. [[CrossRef](#)]
- Zhang, Y.; Wang, X.; Han, N.; Zhao, R. Ensemble Learning Based Postpartum Hemorrhage Diagnosis for 5G Remote Healthcare. *IEEE Access* **2021**, *9*, 18538–18548. [[CrossRef](#)]
- Begum, M.; Redoy, R.M.; Anty, A.D. Preterm Baby Birth Prediction using Machine Learning Techniques. In Proceedings of the 2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD), Dhaka, Bangladesh, 27–28 February 2021; pp. 50–54. [[CrossRef](#)]
- Moreira, M.W.L.; Rodrigues, J.J.P.C.; Al-Muhtadi, J.; Korotaev, V.V.; de Albuquerque, V.H.C. Neuro-fuzzy model for HELLP syndrome prediction in mobile cloud computing environments. *Concurr. Comput. Pract. Exp.* **2021**, *33*, e4651. [[CrossRef](#)]

8. Lin, M.; He, X.; Guo, H.; He, M.; Zhang, L.; Xian, J.; Lei, T.; Xu, Q.; Zheng, J.; Feng, J.; et al. Use of real-time artificial intelligence in detection of abnormal image patterns in standard sonographic reference planes in screening for fetal intracranial malformations. *Ultrasound Obs. Gynecol.* **2022**, *59*, 304–316. [[CrossRef](#)]
9. Mol, B.W.J.; Roberts, C.T.; Thangaratinam, S.; Magee, L.A.; de Groot, C.J.M.; Hofmeyr, G.J. Pre-eclampsia. *Lancet* **2016**, *387*, 999–1011. [[CrossRef](#)]
10. Dadelszen, P.V.; Magee, L.A.; Roberts, J.M. Subclassification of preeclampsia. *Hypertens. Pregnancy* **2003**, *22*, 143–148. [[CrossRef](#)]
11. García, I.H.; Jiménez, A.E.L.; Arriaga, P.I.G.; Abad, D.E.; Izquierdo, A.G. Doppler de arterias uterinas y marcadores angiogénicos (sFlt-1/PIGF): Futuras implicaciones para la predicción y el diagnóstico de la preeclampsia. *Diagn. Prenat.* **2011**, *22*, 32–40. [[CrossRef](#)]
12. Fetal Growth Restriction: ACOG Practice Bulletin, Number 227. *Obstet. Gynecol.* **2021**, *137*, e16–e28. [[CrossRef](#)]
13. Sharma, D.; Shastri, S.; Sharma, P. Intrauterine Growth Restriction: Antenatal and Postnatal Aspects. *Clin. Med. Insights. Pediatr.* **2016**, *10*, 67. [[CrossRef](#)] [[PubMed](#)]
14. Friedman, A.M.; Cleary, K.L. Prediction and prevention of ischemic placental disease. *Semin. Perinatol.* **2014**, *38*, 177–182. [[CrossRef](#)] [[PubMed](#)]
15. Burton, G.J.; Redman, C.W.; Roberts, J.M.; Moffett, A. Pre-eclampsia: Pathophysiology and clinical implications. *BMJ* **2019**, *366*, L2381. [[CrossRef](#)] [[PubMed](#)]
16. Sufriyana, H.; Wu, Y.W.; Su, E.C.Y. Prediction of Preeclampsia and Intrauterine Growth Restriction: Development of Machine Learning Models on a Prospective Cohort. *JMIR Med. Inform.* **2020**, *8*, e15411. [[CrossRef](#)] [[PubMed](#)]
17. Luo, W.; Phung, D.; Tran, T.; Gupta, S.; Rana, S.; Karmakar, C.; Shilton, A.; Yearwood, J.; Dimitrova, N.; Ho, T.B.; et al. Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research: A Multidisciplinary View. *J. Med. Internet Res.* **2016**, *18*, e323. [[CrossRef](#)] [[PubMed](#)]
18. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
19. Premru-Srsen, T.; Premru-Srsen, T.T. Uterine arteries Doppler and sFlt-1/PIGF ratio in hypertensive disorders during pregnancy. *Mendeley Data* **2018**, *V1*. [[CrossRef](#)]
20. Fabjan-Vodusek, V.; Kumer, K.; Osredkar, J.; Verdenik, I.; Gersak, K.; Premru-Srsen, T. Correlation between uterine artery Doppler and the sFlt-1/PIGF ratio in different phenotypes of placental dysfunction. *Hypertens. Pregnancy* **2019**, *38*, 32–40. [[CrossRef](#)] [[PubMed](#)]
21. Kuhn, M.; Johnson, K. *Applied Predictive Modeling*; Springer Science & Business Media: New York, NY, USA, 2013. [[CrossRef](#)]
22. Herrera, F.; Charte, F.; Rivera, A.J.; del Jesus, M.J. Multilabel Classification. In *Multilabel Classification: Problem Analysis, Metrics and Techniques*; Springer: Cham, Switzerland, 2016; pp. 17–31. [[CrossRef](#)]
23. Luaces, O.; Díez, J.; Barranquero, J.; del Coz, J.J.; Bahamonde, A. Binary relevance efficacy for multilabel classification. *Prog. Artif. Intell.* **2012**, *1*, 303–313. [[CrossRef](#)]
24. Boutell, M.R.; Luo, J.; Shen, X.; Brown, C.M. Learning multi-label scene classification. *Pattern Recognit.* **2004**, *37*, 1757–1771. [[CrossRef](#)]
25. Sibai, B.; Dekker, G.; Kupferminc, M. Pre-eclampsia: A first hand account. *Lancet* **2005**, *365*, 785–799. [[CrossRef](#)]
26. Duley, L. The global impact of pre-eclampsia and eclampsia. *Semin. Perinatol.* **2009**, *33*, 130–137. [[CrossRef](#)] [[PubMed](#)]
27. Allotey, J.; Snell, K.I.E.; Smuk, M.; Hooper, R.; Chan, C.L.; Ahmed, A.; Chappello, L.C.; von Dadelszen, P.; Dodds, J.; Green, M.; et al. Validation and development of models using clinical, biochemical and ultrasound markers for predicting pre-eclampsia: An individual participant data meta-analysis. *Health Technol. Assess* **2020**, *24*. [[CrossRef](#)] [[PubMed](#)]
28. Poon, L.C.; Shennan, A.; Hyett, J.A.; Kapur, A.; Hadar, E.; Divakar, H.; McAuliffe, F.; de Costa Silva, F.; von Dadelszen, P.; McIntyre, H.D.; et al. The International Federation of Gynecology and Obstetrics (FIGO) initiative on pre-eclampsia: A pragmatic guide for first-trimester screening and prevention. *Int. J. Gynecol. Obs.* **2019**, *145*, 1–33. [[CrossRef](#)]
29. ACOG committee opinion no. 743: Low-dose aspirin use during pregnancy. *Obstet. Gynecol.* **2018**, *132*, e44–e52. [[CrossRef](#)]
30. Roberge, S.; Nicolaides, K.; Demers, S.; Hyett, J.; Chaillet, N.; Bujold, E. The role of aspirin dose on the prevention of preeclampsia and fetal growth restriction: Systematic review and meta-analysis. *Am. J. Obstet. Gynecol.* **2017**, *216*, 110–120.e6. [[CrossRef](#)]
31. Rolnik, D.L.; Wright, D.; Poon, L.C.; O’Gorman, N.; Syngelaki, A.; de Paco Matallana, C.; Akolekar, R.; Cicero, S.; Janga, D.; Singh, M.; et al. Aspirin versus Placebo in Pregnancies at High Risk for Preterm Preeclampsia. *N. Engl. J. Med.* **2017**, *377*, 613–622. [[CrossRef](#)] [[PubMed](#)]
32. Moller, A.B.; Petzold, M.; Chou, D.; Say, L. Early antenatal care visit: A systematic analysis of regional and global levels and trends of coverage from 1990 to 2013. *Lancet Glob. Health* **2017**, *5*, e977–e983. [[CrossRef](#)]