PRELIMINARY RESULTS OF AI APPLICATION IN PREDICTING PRENATAL WEIGHT IN VIET NAM

**Pham Ba Nha[1], Nguyen Phuong Tu[2,3], Tran Le Duc Anh[4], Do Ba Quang Huy[5],
Nguyen Hai Phuong[6], Nguyen Anh Thu[7],Nguyen Phuong Hong Minh[7],
Le Thien Thai[1], Nguyen Thi Huong Linh[1], Nguyen Tien Hoang[1]**

ABTRACT

Objective: Our research aims to propose the application of AI in improving the accuracy for estimating fetal weight.

Method: This is a cross-sectional analytical study combining the development and comparison of predictive models supported by AI that include 1162 pregnancy women that came for scheduled visit and give birth at Obstetric Department at Vinmec Times City International Hospital. The pregnancy women were chosen based on: age (18 to 35 years old), live singleton pregnancy, Vietnamese, height > 153 cm, and gestational age ≥ 36 weeks. Maternal and fetal parameters are then collected and divided into data and test set. Symbolic regression learning method is applied to our ensemble model which include Operon, XGBoost, LightGBM, FEAT, Hadlock, GP-GOMEA, DSR, and SBP-GP. The performance of each algorithms are then evaluated using test data set and compared the results to the traditional method.

Results: 1162 pregnancy women were selected, and each were collected 9 maternal and fetal variables. MAE, MSE, RMSE, and R^2 value of each model were calculated from their performance. Operon model's MAE was 158.542 grams, RMSE was 205.485 and R^2 of 0.538, while Hadlock model's MAE was 186.98 grams, RMSE was 331.373 and R^2 of -0.19.

Conclusion: AI-driven methods, particularly symbolic regression, demonstrates greater accuracy and efficacy compared to the traditional method, hence suggests a promising potential in prenatal weight estimation.

Keywords: symbolic regression, fetal weight estimation, ensemble learning, machine learning

[1]*Women's Health Center, Vinmec Times City Hospital*

[2]*Ha Noi Medical University*

[3]*Bach Mai Hospital*

[4]*School of Electrical and Electronic Engineering, Hanoi University of Science and Technology*

[5]*FPT automative, FPT Software*

[6]*National hospital of obstetrics and gynecology*

[7]*College of Health Sciences, Vin University*

Responsible person: Nguyen Tien Hoang
Email: nguyentienhoangyds14@gmail.com
Date of receipt: 10/2/2025
Date of scientific judgment: 10/3/2025
Reviewed date: 17/3/2025

I. INTRODUCTION

In obstetrics, ultrasound has a crucial role in monitoring fetal growth by providing some important parameters, such as bitemporal diameter (BDP), head circumference (HC), abdominal circumference (AC), and femoral length (FL). Predicting fetal birth weight allows medical professionals to evaluate the possibility of preterm birth and decide method of delivery. Additionally, estimating fetal weight helps to detect the intrauterine growth restrictions or large for gestational age, both of which can lead to serious complications such as stillbirth or pregnancy loss. Therefore, it is necessary to improve the accuracy of EFW in an ultrasound's formula.

To estimate the fetal weight, we applied the use of clinical-based method, either the classic method of measuring maternal abdominal circumference or using formula to calculate based on fetal biometrics measured

by ultrasound, and most commonly used were the Hadlock's formula (1). According to Nguyen et al., his research on more than 500 pregnancy women showed that the difference in efficacy of both the classic method and the ultrasound-based method was unsignificant, and this point of view was mentioned by multiple researchers (2). However, despite the widely application of both methods, they were having one common variable – human. The difference in each practitioners' observations, experience, and method of application results in the difference when measuring each parameters, hence eventually leading the biggest limitations of both methods – wide range of deviations compared to the actual birth weight (3).

Nowadays, Hadlock IV and INTERGROWTH-21 (IG-21) are two formulas used as standard growth charts in fetal ultrasound in many countries around the world (1). Several studies have demonstrated that Hadlock and IG-21 have high accuracy in determining fetal weight, although there is still some debate about which formula is more optimal (4). However, research results in Asian populations have shown that the Hadlock V formula is not optimal in estimating fetal birth weight, especially in LGA or IUGR populations (5), particularly in East and Southeast Asian countries In Vietnam, in the LGA group, the error in weight estimation is higher compared to other groups, or in some fetal groups, predicting fetal weight using ultrasound formulas is not as good as using classical formulas through clinical examination (2). Malaysia also reported that applying Hadlock

or IG-21 can lead to misdiagnosis of small for gestational age (SGA) and that using their own growth chart is even more accurate (6). This has emphasized the urgent need for a nationalized fetal weight estimation formula that is more suitable for the Vietnamese population.

Recently, AI has been used in multiple aspects of daily life, and particularly in medical field. In the obstetrics-gynecology (OBGYN) ultrasound, several researches has been conducted to report the application of AI, such as in estimating placental volume, identifying lesions on colposcopy image, measuring endometrial thickness and classify ovarian cyst, and automatically detect the planes in fetal biometrics ultrasound. In order to do so, AI has been using Deep Learning method through deep convolutional neural network. If successfully applying AI in estimating fetal weight, it would help reducing time for each scan, also able to avoid subjective errors from the technicians, which all serves the purpose to improve the accuracy of ultrasound fetal biometric measurements. With an idea of creating a model supported by AI to efficiently estimate pre-labor fetal weight for the Vietnamese population.

II. RESEARCH OJECTIVE AND METHOD
2.1. Research design
The research is a cross-sectional analytical study combining the development and comparison of predictive models supported by AI.
2.2. Research object
In this study, there are selected 1162 pregnant women attending for check-ups and

pregnancy monitoring and give birth at the Obstetrics Department of Vinmec Times City Hospital meet the following inclusion criteria and Exclusion criteria:

Inclusion criteria:

• Pregnant women aged between 18 and 35 years.

• Singleton live pregnancy.

• Both the pregnant woman and her spouse are Vietnamese nationals.

• The pregnant woman's height is 153 cm or above.

• Gestational age: From 36 weeks until delivery, for managing and evaluating ultrasound-measured weight and actual birth weight (if multiple ultrasounds are performed from 36 weeks to delivery, the most recent ultrasound result will be used).

Exclusion criteria:

• Pregnancy with congenital anomalies, stillbirth, or intrauterine growth restriction.

• Refusal to participate in the study or withdrawal from the study.

2.3. Dataset detail

The dataset comprises 1162 observations and 9 variables, collected from patients visiting Vinmec Hospital for prenatal examinations. It includes essential maternal and fetal health parameters such as *Gestational age* (measured in weeks), *Weight* (in grams), and *Gestational age at last ultrasound*. Detailed fetal measurements, including *Biparietal diameter (BPD)*, *Head Circumference (HC)*, *Waist circumference (AC)*, and *Femur length (FL),* are provided in millimeters. The dataset also contains a binary variable, *T/G*, which categorizes patients based on a specific condition, and *Mother Age* in years. For example, one record shows a 30-year-old mother delivering

a baby at 39.7 weeks of gestation, with a birth weight of 2840 grams. This dataset offers a rich resource for investigating the relationships between maternal characteristics, fetal growth patterns, and birth outcomes, with data rooted in a real-world clinical context at Vinmec Hospital.

2.4. Research method

We evaluated the dataset from Vinmec Hospital on two distinct tasks. First, we assessed the capability of various models to predict neonatal birth outcomes (e.g., birth weight) using maternal and fetal characteristics while minimizing the complexity of the resulting models. Second, we evaluated the models' ability to accurately capture relationships between variables, such as gestational age, fetal measurements, and maternal age, under different conditions.

The experimental settings are summarized in Figure 1. Each algorithm was trained on the dataset using a 10-fold cross-validation approach, repeated 10 times with different random seeds controlling the train/test split and the algorithm initialization. The dataset was divided into a 75/25% training and testing split. For the predictive tasks, hyperparameter tuning was performed using 5-fold cross-validation with halving grid search. The models were restricted to six combinations of hyperparameters to ensure fair comparison. Evaluation metrics included mean squared error (RMSE) and R-squared for regression tasks, and accuracy for classification tasks based on the binary T/G variable. Further details on the experimental setup, including noise levels for robustness testing and implementation specifics, can be found in below.
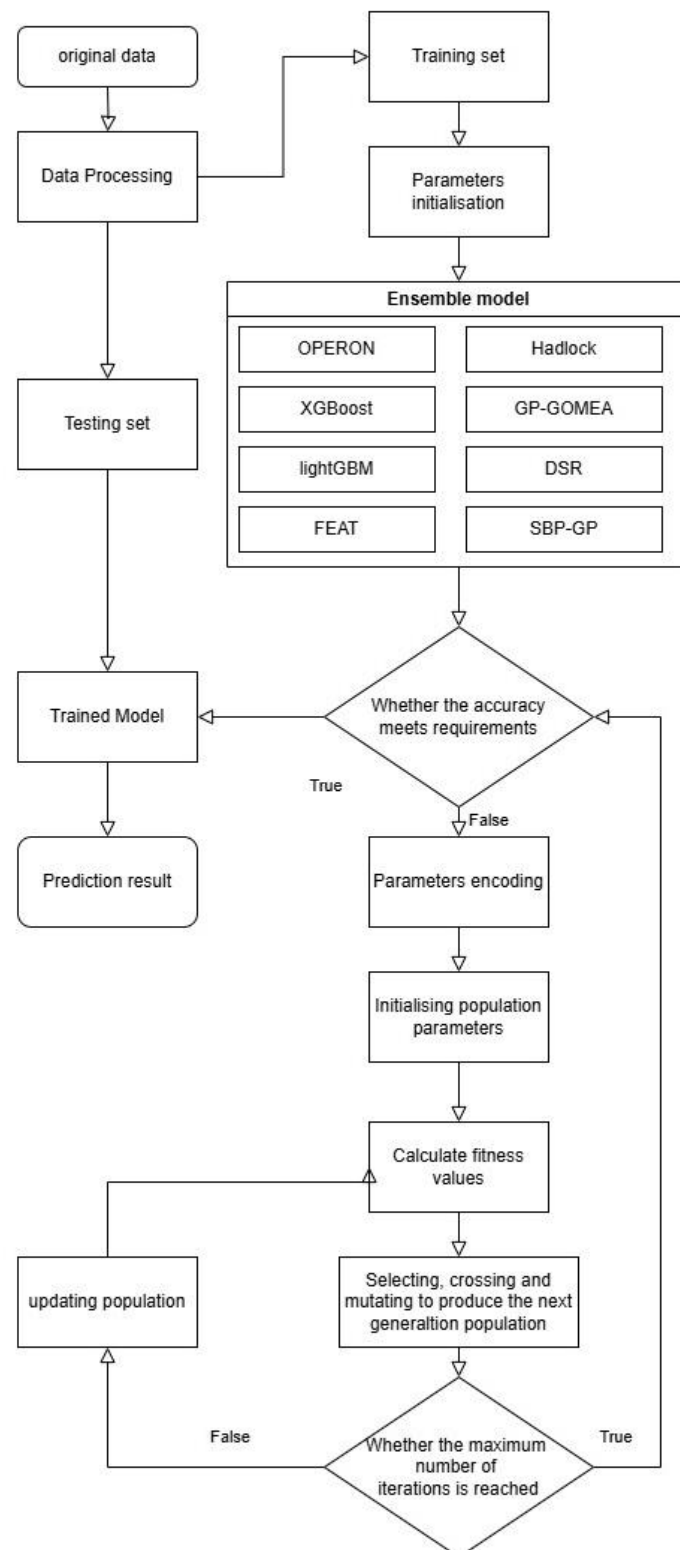
Figure 1: **The flowchart of the genetic algorithm**

Algorithm Details

Here, we characterize the symbolic regression (SR) methods applied in our study, emphasizing how they align with broader trends in SR research and their specific applicability to the clinical dataset. The SR methods used span traditional genetic programming approaches, modern gradient-based techniques, and Bayesian optimization methods.

Operon

$$f(x) = c_1.f_1(x) + c_2.f_2(x) + c_3.f_3(x) + ... + c_n.f_n(x)$$

where $f_i(x)$ are basis functions evolved during training, and c_i are constantly optimized using non-linear least squares. By focusing on improving numerical constants, Operon achieves superior performance in noisy datasets while maintaining model interpretability.

Semantic Backpropagation Genetic Programming (SBP-GP)

SBP-GP enhances traditional GP by using semantic information to guide model optimization. During training, intermediate outputs from symbolic expressions are compared to the target outputs, and values are adjusted dynamically to align predictions with ground truth. This semantic backpropagation approach reduces premature convergence and mitigates overfitting, particularly in datasets with high noise levels. Virgolin et al. demonstrated the utility of SBP-GP in applications where intermediate corrections to symbolic expressions yield significant improvements in accuracy and simplicity.

Feature Engineering Automation Tool (FEAT)

FEAT combines symbolic regression with feature engineering to produce interpretable, high-performance models. It evolves interpretable features represented as mathematical expressions:

Operon is a symbolic regression framework that combines genetic programming (GP) with advanced optimization techniques to generate interpretable mathematical models. Operon utilizes the **Levenberg-Marquardt algorithm** for optimizing constants within symbolic expressions, ensuring both accuracy and parsimony. The method optimizes equations of the form:

interpretable features represented as mathematical expressions:

$$f(x) = \beta_0 + \sum_{i=1}^{k} \beta_1. \phi_i(x)$$

where $\phi_i(x)$ evolved symbolic features and β_1 are coefficients optimized via gradient descent. Each evolved model in FEAT represents a linear combination of features, where the features themselves are mathematical expressions generated through GP. The parameters of these features are further refined using gradient descent, ensuring that the final models are both accurate and easy to interpret. FEAT's ability to automate feature discovery while maintaining domain-specific constraints has been highlighted as a breakthrough in SR research, particularly for applications requiring human interpretability, such as healthcare analytics.

GP-GOMEA

The Gene-pool Optimal Mixing Evolutionary Algorithm for Genetic Programming (GP-GOMEA) adapts traditional GP by incorporating dependency-aware recombination. In GP-GOMEA, statistical models of variable dependencies are constructed during each generation.

These models guide the recombination of components within symbolic equations, ensuring that interdependent subcomponents are preserved. This approach has been shown to accelerate convergence and reduce bloating, making it well-suited for datasets with complex, non-linear relationships.

Deep Symbolic Regression (DSR)

DSR is a reinforcement learning-based approach to SR that uses a recurrent neural network (RNN) to generate symbolic expressions. The network is trained to maximize a reward function:

$$R = -Loss + \lambda.Parsimony$$

The expressions are evaluated using a reward function that balances prediction accuracy with parsimony, ensuring that generated models remain interpretable. DSR employs a novel variant of the Monte Carlo policy gradient algorithm, dubbed "risk-seeking policy gradient," which biases the training process towards discovering exact mathematical solutions. This makes DSR particularly effective for tasks where discovering the underlying functional form is critical, as in synthetic datasets or controlled experimental data.

Hadlock Algorithm

The Hadlock method, a standard in obstetric medicine, estimates fetal weight using ultrasound measurements. It incorporates parameters such as biparietal diameter (BPD), head circumference (HC), abdominal circumference (AC), and femur length (FL) into regression equations validated across diverse populations. By integrating these clinical formulas into SR models, we establish a benchmark for evaluating the performance of modern machine learning and SR techniques in predicting birth outcomes. The Hadlock algorithm's robustness and interpretability make it a valuable tool for clinical decision-making (1). In this paper we use all Hadlock I, Hadlock II, Hadlock III, Hadlock IV and we take the best results of it to represent for Hadlock.

III. RESULT

To evaluate the performance of the models comprehensively, we used the following metrics: coefficient of determination (R^2) and root mean squared error (RMSE). Each metric captures a different aspect of model performance, such as goodness of fitness, accuracy, or error magnitudes. The R^2 measures the proportion of the variance in the dependent variable (yyy) that is predictable from the independent variables (xxx). RMSE quantifies the standard deviation of the prediction errors and emphasizes larger errors more than MAE due to squaring the residuals. RMSE is sensitive to outliers, making it a useful metric when large prediction errors are particularly undesirable.

Table 1: **MAE, MSE, RMSE and R² of different models**

| Model	MAE	MSE	RMSE	R^2
Operon	158.542	42224.49	205.485	0.538
XG-Boost	169.934	49896.06	212.419	0.504
Light-GBM	166.543	40302.31	172.041	0.489
FEAT	174.715	38061.05	183.200	0.623
Hadlock IV	186.98	109802.85	331.373	-0.19
GP-GOMEA	235.690	49797.920	207.565	0.488
DSR	191.804	45292.658	202.769	0.453
SBP-GP	189.146	43205.621	218.399	0.408

In this study, we focused on two representative examples: Hadlock IV, a traditional method, and Operon, an AI-based method using symbolic regression, to highlight the advantages of applying modern machine learning techniques over conventional approaches. While all tested models provided valuable insights, we chose to display only the results of these two methods for clarity and emphasis, as the outcomes from the other models (such as XG-Boost, Light-GBM, and FEAT) were similarly consistent with the trend observed.

Hadlock IV, a well-known method in clinical practice, is based on established algorithms for estimating fetal weight using biometric measurements from ultrasound scans. While this method has been widely used in obstetrics for many years, its predictive accuracy (as shown by the higher MAE, MSE, and RMSE values) is often limited by its reliance on a set of predefined factors and assumptions, which may not always capture the complexity and variability of pregnancy outcomes. In this study, Hadlock IV produced a MAE of 186.98 grams, an MSE of 109802.85, and an RMSE of 331.37 grams, with a negative R^2 value (-0.19), suggesting poor fit for the dataset.

On the other hand, Operon, an AI-driven method based on symbolic regression, demonstrated superior accuracy. Symbolic regression allows the model to explore a wide range of mathematical expressions and relationships in the data, thus leading to better adaptability and predictive power. As evidenced in the results, Operon achieved a MAE of 158.54 grams, an MSE of 42224.49, and an RMSE of 205.49 grams, with a positive R^2 of 0.538, indicating a much better fit to the data. This impressive performance highlights the strength of AI models in capturing complex patterns that traditional methods struggle to identify.
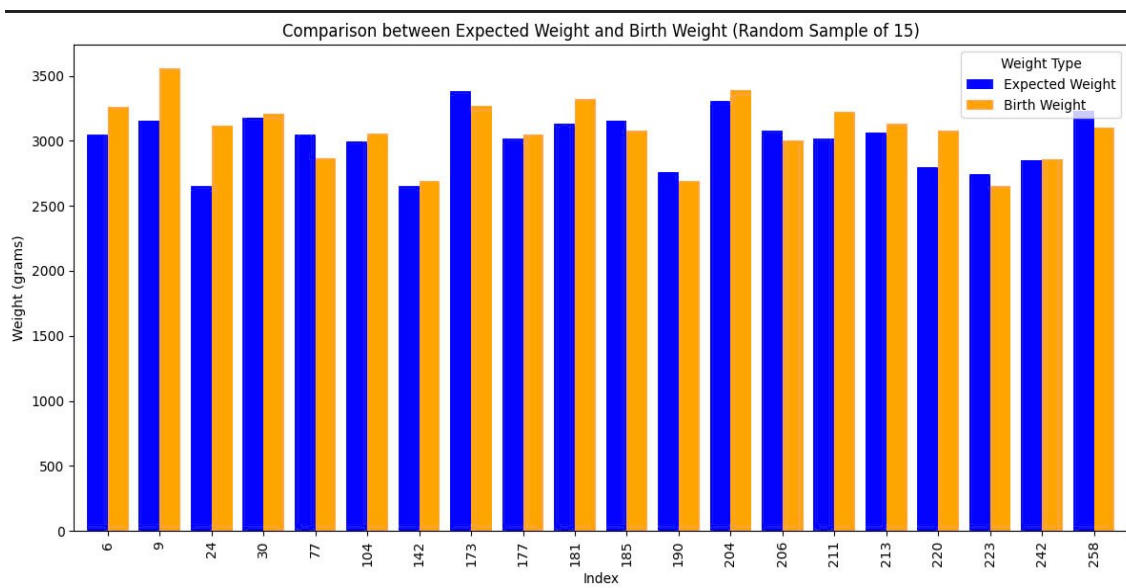
Figure 2: **Comparison between Expected Weight and Birth Weight in Operon Model (Random Sample of 15)**

The decision to showcase Operon alongside Hadlock IV serves to underscore the contrast between a traditional model and a modern machine learning approach. The consistent trend observed in the performance metrics of the other models suggests that they would likely produce similar results to Hadlock IV, demonstrating the clear

advantage of AI-based models, like Operon, in making more accurate predictions.

This comparison further illustrates how symbolic regression and AI-driven approaches can offer more precise, adaptable, and efficient solutions for healthcare tasks, particularly in applications such as predicting fetal birth weight, where accuracy is critical. While traditional methods like Hadlock IV remain valuable and widely used in clinical settings, AI methods like Operon are poised to revolutionize healthcare by enabling more accurate predictions and better decision-making capabilities.

IV. DISCUSSION

Fetal ultrasound is an essential tool for monitoring fetal growth and estimating fetal weight based on formulas that help physicians make decisions about pregnancy trimester management and delivery method. Using data from 1171 patients at Vinmec Hospital, this study used that data to investigate the accuracy of EFW compared to actual birth weight by applying the data to 7 symbolic regression models and 1 popular traditional formula - Hadlock IV. The results showed that AI-driven models (Operon, FEAT, GP-GOMEA, DSR) generally outperformed traditional models (Hadlock IV) in predicting fetal birth weight.

The research findings indicate that AI-based predictive models demonstrate higher accuracy compared to traditional models. This viewpoint is consistent with Lu et al., who developed an ensemble model using Random Forest, XGBoost, and LightGBM and found that their model surpassed traditional methods in predicting fetal weight in the absence of ultrasound (8). This convergence of findings across different methodologies and datasets persuades the argument for integrating advanced machine learning techniques into clinical practice. In this study, Operon, which represents AI-driven symbolic regression models, is compared to Hadlock IV because other modern models also share the same trend. Operon presents the closer to the actual birth weight (lower MAE), significant reduction in prediction error variability (MSE), improve accuracy (RMSE) & superior fit (R^2). The negative R^2 of Hadlock IV suggests that conventional, pre-established formulas fail to fully capture the diversity present in biological data. In comparison, the implication of our study is the potential of AI-driven models to improve the accuracy of fetal ultrasound-based EFW formulas, which has been a limitation of traditional formulas.

Using a real-world dataset from Vinmec Hospital is one of its strongest strengths; it increases the findings' clinical applicability and robustness. A thorough assessment of the models' performance was guaranteed by the 10-fold cross-validation technique, which was carried out ten times using various random seeds. However, the dataset is restricted to a single institution, which could limit how broadly our findings can be applied. Future studies should examine how these models function with broader, more varied datasets and in various clinical contexts. In addition, the models can only predict fetal weight from 36 weeks gestation under specific conditions, such as a singleton pregnancy, maternal age between 18 and 35 years, maternal height above 153 cm, no complications during pregnancy, and the absence of fetal abnormalities. Therefore, it cannot be applied to predict fetal weight at multiple time points throughout pregnancy. However, the research findings suggest that AI-based deep learning models have significant potential in predicting fetal weight compared to traditional models.

While symbolic regression is well-known for its interpretability and accuracy, it is still not widely applied in the field, especially in estimating fetal weights, where the deep

learning models take over the dominance. However, despite the methods being used, we all came to a consensus that machine learning would outweighs conventional approaches in terms of accuracy. Hence, in the near future, we hope the symbolic regression method would be able to emphasize its function by either combining with other AI-driven method, or by applying it to a greater data size.

From our research, we believe that, with the use of machine learning, especially symbolic regression method, it would provide a great potential to the field of estimating fetal weights, where accuracy and timely decision are utmost priority. And when the estimated fetal weights are more precise, it would eventually improve the pregnancy outcome by able to provide appropriate interventions and lowering the complication rate during labour. The potential of AI-driven methods in healthcare are largely untapped, hence various avenues for future research with promising outcomes.

V. CONCLUSION

According to the historical data of the physical examinations of pregnant women, this study uses, the study compared the accuracy of predicting prenatal weight based on fetal BPD, AC, HC, and FL measurements between traditional models and AI-supported models. The research findings show that AI-based predictive models exhibit higher accuracy than traditional models, with RMSE and R^2 values of 205.485 and 0.538, respectively. The study suggests that future research could implement AI-based predictive models on a larger and more diverse sample size to to improve the prediction model to further improve the accuracy and practicability of model's prediction.

REFERENCES

1. **Hadlock FP, Harrist RB, Sharman RS, Deter RL, Park SK.** Estimation of fetal weight with the use of head, body, and femur measurements—A prospective study. Am J Obstet Gynecol. 1985 Feb 1;151(3):333–7.
2. **Nguyen XC, Doan QH, Bui TT, Le NQH, Nguyen QM, Dam LC, et al.** Khảo sát giá trị các phương pháp ước lượng cân nặng thai nhi đủ tháng | Tạp chí Phụ sản. 2023 Sep 23 [cited 2024 Dec 7];21(3). Available from: https://vjog.vn/journal/article/view/1615
3. **Kumari A, Goswami S, Mukherjee P.** Comparative Study of Various Methods of Fetal Weight Estimation in Term Pregnancy. J South Asian Fed Obstet Gynaecol. 2013 Apr;5(1):22–5.
4. **Marien M, Perron S, Bergeron AM, Singbo N, Demers S.** Comparison of the Accuracy of INTERGROWTH-21 and Hadlock Ultrasound Formulae for Fetal Weight Prediction. J Obstet Gynaecol Can JOGC J Obstet Gynecol Can JOGC. 2021 Nov;43(11):1254–9.
5. **Ma J, Cheng D, Zhang Z, Cai B, Xu X.** Evaluating the accuracy of sonographic fetal weight estimations using the Hadlock IV formula in a Chinese population. Quant Imaging Med Surg. 2023 Jun 1;13(6):3726–34.
6. **Saw SN, Lim MC, Liew CN, Ahmad Kamar A, Sulaiman S, Saaid R, et al.** The accuracy of international and national fetal growth charts in detecting small-for-gestational-age infants using the Lambda-Mu-Sigma method. Front Surg [Internet]. 2023 Apr 11 [cited 2024 Nov 26];10. Available from: https://www.frontiersin.org/journals/surgery/articles/10.3389/fsurg.2023.1123948/full
7. **Schmidt M, Lipson H.** Distilling Free-Form Natural Laws from Experimental Data. Science. 2009 Apr 3;324(5923):81–5.
8. **Lu Y, Fu X, Chen F, Wong KKL.** Prediction of fetal weight at varying gestational age in the absence of ultrasound examination using ensemble learning. Artif Intell Med. 2020 Jan 1;102:101748.