BMC Pregnancy and Childbirth

RESEARCH

Open Access

Predictive modeling of gestational weight gain: a machine learning multiclass classification study

Audêncio Victor[1*], Hellen Geremias dos Santos[2], Gabriel Ferreira Santos Silva[1], Fabiano Barcellos Filho[1], Alexandre de Fátima Cobre[3], Liania A. Luzia[1], Patrícia H.C. Rondó[1] and Alexandre Dias Porto Chiavegatto Filho[1]

Abstract

Background Gestational weight gain (GWG) is a critical factor influencing maternal and fetal health. Excessive or insufficient GWG can lead to various complications, including gestational diabetes, hypertension, cesarean delivery, low birth weight, and preterm birth. This study aims to develop and evaluate machine learning models to predict GWG categories: below, within, or above recommended guidelines.

Methods We analyzed data from the Araraquara Cohort, Brazil, which comprised 1557 pregnant women with a gestational age of 19 weeks or less. Predictors included socioeconomic, demographic, lifestyle, morbidity, and anthropometric factors. Five machine learning algorithms (Random Forest, LightGBM, AdaBoost, CatBoost, and XGBoost) were employed for model development. The models were trained and evaluated using a multiclass classification approach. Model performance was assessed using metrics such as area under the ROC curve (AUC-ROC), F1 score and Matthew's correlation coefficient (MCC).

Results The outcomes were categorized as follows: GWG within recommendations (28.7%), GWG below (32.5%), and GWG above recommendations (38.7%). The XGBoost presented the best overall model, achieving an AUC-ROC of 0.79 for GWG within, 0.76 for GWG below, and 0.65 for GWG above. The LightGBM also performed well with an AUC-ROC of 0.79 for predicting GWG within recommendations, 0.76 for GWG below, and 0.624 for GWG above. The most important predictors of GWG were pre-gestational BMI, maternal age, glycemic profile, hemoglobin levels, and arm circumference.

Conclusion Machine learning models can effectively predict GWG categories, offering a valuable tool for early identification of at-risk pregnancies. This approach can enhance personalized prenatal care and interventions to promote optimal pregnancy outcomes.

Keywords Gestational weight gain, Machine learning, Prediction models, Maternal health, Fetal health, Araraquara cohort

*Correspondence:
Audêncio Victor
audenciovictor@gmail.com
[1]School of Public Health, University of São Paulo (USP), Avenida Doutor Arnaldo, 715, São Paulo 01246904, São Paulo, Brazil
[2]Oswaldo Cruz Foundation - Carlos Chagas Institute (ICC) Parana, Curitiba, Paraná, Brazil
[3]Department of Statistics, Federal University of Paraná, Curitiba, Paraná, Brazil

Introduction

Gestational weight gain (GWG) has been shown to directly influence maternal and infant health outcomes [1–3]. Excessive GWG is associated with gestational complications, such as gestational diabetes and hypertension, as well as long-term risks like metabolic and cardiometabolic diseases in childhood [4, 5]. On the other hand, insufficient GWG increases the risk of low birth weight, preterm birth, and perinatal mortality [6–8]. On the other hand, insufficient GWG increases the risk of low birth weight, preterm birth, and perinatal mortality. These outcomes are further affected by factors such as pre-pregnancy body mass index (BMI), maternal age, sociodemographic conditions, and race [9, 10].

Machine learning (ML), a subfield of artificial intelligence (AI), offers new opportunities for analyzing large volumes of data (big data) and identifying complex patterns that traditional statistical methods may not capture [11–13]. The application of ML techniques in public health has rapidly expanded, providing powerful tools for prediction, diagnosis, and monitoring of health conditions [14–17]. In the context of perinatal health, accurate prediction of GWG can enable the early identification of at-risk pregnant women and the implementation of targeted interventions.

Despite the promising potential of ML, the literature remains scarce in basic or translational research that uses AI to predict maternal and infant outcomes, especially in low-income regions and with limited sample sizes [16, 18–20]. Recent data indicate that a significant proportion of pregnant women do not meet the recommended parameters set by the Institute of Medicine (IOM), highlighting the need for personalized and early interventions to improve pregnancy outcomes [21, 22]. This study aims to fill this gap by applying advanced ML techniques to predict categories of GWG. The objective of this study is to identify women at higher risk of inadequate weight gain during pregnancy, enabling preventive interventions that promote healthy pregnancy outcomes. Using longitudinal data from the Araraquara Cohort, we tested and compared the performance of ML algorithms in a multiclass classification approach. Our results aim to contribute to the improvement of personalized prenatal care and the reduction of disparities in maternal and infant health outcomes.

Materials and methods

Dataset description

We analyzed data from a population-based cohort study conducted in Araraquara, São Paulo, Brazil (Araraquara Cohort). The sample included women with a gestational age less than or equal to 19 weeks, who received prenatal care at Basic Health Units in Araraquara. Participants were followed quarterly throughout prenatal care until the birth of their children from 2017 to 2022. Excluded from the study were women with twin pregnancies and those who had a pre-viable abortion. In cases of fetal death and stillbirths, only pregnancy data were considered.

Several characteristics were considered for predicting GWG as shown in Fig. 1. Socioeconomic and demographic factors included age (≤19, 20–35, or >35 years), educational level (<4, 5–11, or ≥12 years of schooling), per capita income in Brazilian reais (1 US$ = 4.9 R$), race (white or non-white), marital status (married/stable union or single/separated/widowed), and the number of previous pregnancies (0, 1, or ≥2). Lifestyle factors included physical activity (was assessed using the International Physical Activity Questionnaire, a widely validated tool that measures the frequency, intensity, and duration of physical activity), smoking, and alcohol consumption. Morbidity factors included pre-pregnancy conditions such as diabetes and hypertension, as well as urinary tract infection and cervicitis/vaginitis. Anthropometric data of the pregnant women were evaluated based on height (cm) categorized into tertiles; BMI (kg/m²); arm circumference (cm); and body fat percentage. Other relevant data included gestational age at birth, glycemic profile (fasting glucose [mg/dL], insulin [μUI/mL], HOMA [μUI/mL], glycated hemoglobin [%]), high-sensitivity C-reactive protein (hs-CRP [ng/mL]), hemoglobin [g/dL], and lipid profile (total cholesterol, LDL-c, HDL-c, and triglycerides [mg/dL]). Additionally, the number of family members per room was categorized into tertiles, and the number of previous pregnancies was categorized as 0, 1, and ≥2.

Outcome definition

GWG was calculated as the difference between weight at delivery and pre-pregnancy weight. GWG was then classified into three categories according to the recommendations of the Institute of Medicine (IOM): (a) GWG below IOM, (b) GWG within IOM, and (c) GWG above IOM [22].

Statistical analysis

Descriptive statistics were used to summarize the characteristics of the study population. Continuous Predictors were presented as median and interquartile range (IQR), while categorical Predictors were presented as frequencies and percentages. Differences between GWG categories were tested using the Kruskal-Wallis test for continuous Predictors and the Chi-square test or Fisher's exact test for categorical Predictors.

Machine learning model design

Considering the different outcomes related to GWG, we employed a multiclass classification approach to evaluate

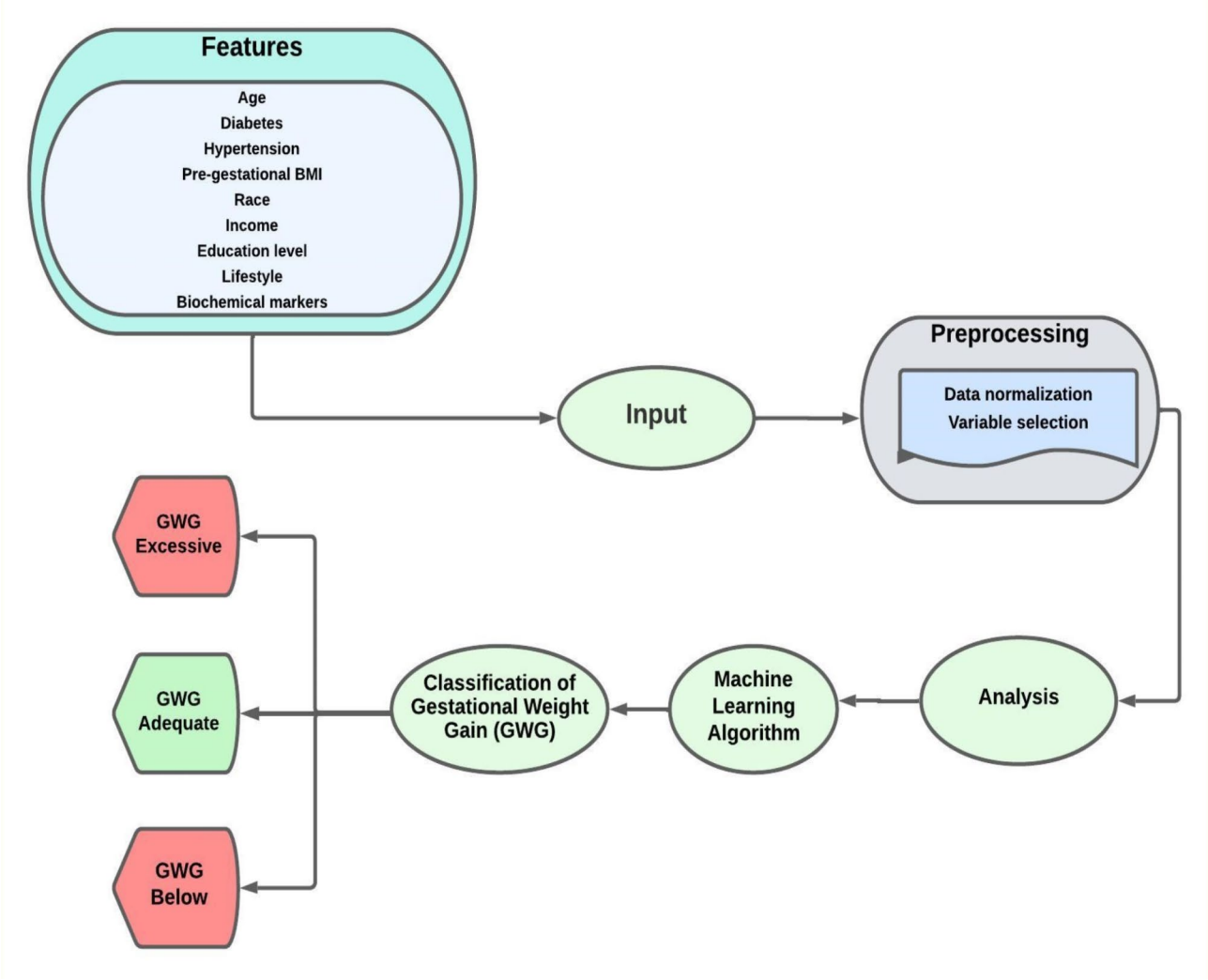

Fig. 1 Workflow diagram for classifying GWG adequacy

whether changing strategies could enhance model performance. Separate models were developed for each GWG category: below IOM recommendations, within IOM recommendations, and above IOM recommendations. The models were evaluated independently without sharing any information during the process, as shown in Fig. 1.

ML techniques

Data preprocessing

For quantitative predictors, standardization was performed using the z-score, separately in the training and test sets. All qualitative predictors were handled through one-hot encoding, where each category was considered separately for this procedure. Additionally, predictors with a percentage of missing values above 20% were removed, while those with less than 20% missing values were imputed using the mean, as recommended by previous studies in healthcare [23, 24].

Algorithm selection

We tested five different ML algorithms: CatBoost [25], XGBoost [26], LightGBM [27], and Random Forest. For CatBoost, XGBoost, and LightGBM, we used their respective Python packages. For the other algorithms, we used the scikit-learn library [28]. Additionally, we employed the bootstrapping technique to further ensure the robustness and reliability of the model's performance.

Hyperparameter selection

Hyperparameter selection in the training set was performed through 10-fold cross-validation, using Bayesian optimization and RandomSearch strategies [29]. In cases of significant class imbalance, where the minority class represented less than 25% of the total outcomes, the Synthetic Minority Over-sampling Technique (SMOTE) was applied. Additionally, in the training set, the BORUTA method was employed for feature selection [30].

Model evaluation

The models with the best performance in the training set (which corresponded to 70% of the data) were selected for evaluation in the test set (30%). The evaluation of machine learning algorithms was conducted in the test set, based on metrics such as area under the ROC curve (AUC-ROC), area under the precision-recall curve (AUC-PR), precision, recall, positive predictive value, negative predictive value, Matthew's correlation coefficient (MCC) and F1 score. Finally, the interpretation and evaluation of each predictors contribution to the outcome were obtained through the calculation of Shapley values [31–33] in the test set. We adhered to the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) guidelines [34].

Results

Maternal characteristics and GWG

The study included 1557 pregnant women, with 28.7% having GWG within the Institute of Medicine (IOM) recommendation, 32.5% below, and 38.7% above, as shown in Fig. 2. The majority of the women, 76.4%, were aged between 20 and 35 years. Additionally, 53.6% were predominantly non-white. Key characteristics associated with GWG categories included pre-gestational BMI, maternal age, glycemic profile, hemoglobin levels, and arm circumference. The prevalence of diabetes and hypertension was significantly higher among women with GWG above the recommendations ($P<0.001$) (Table 1).

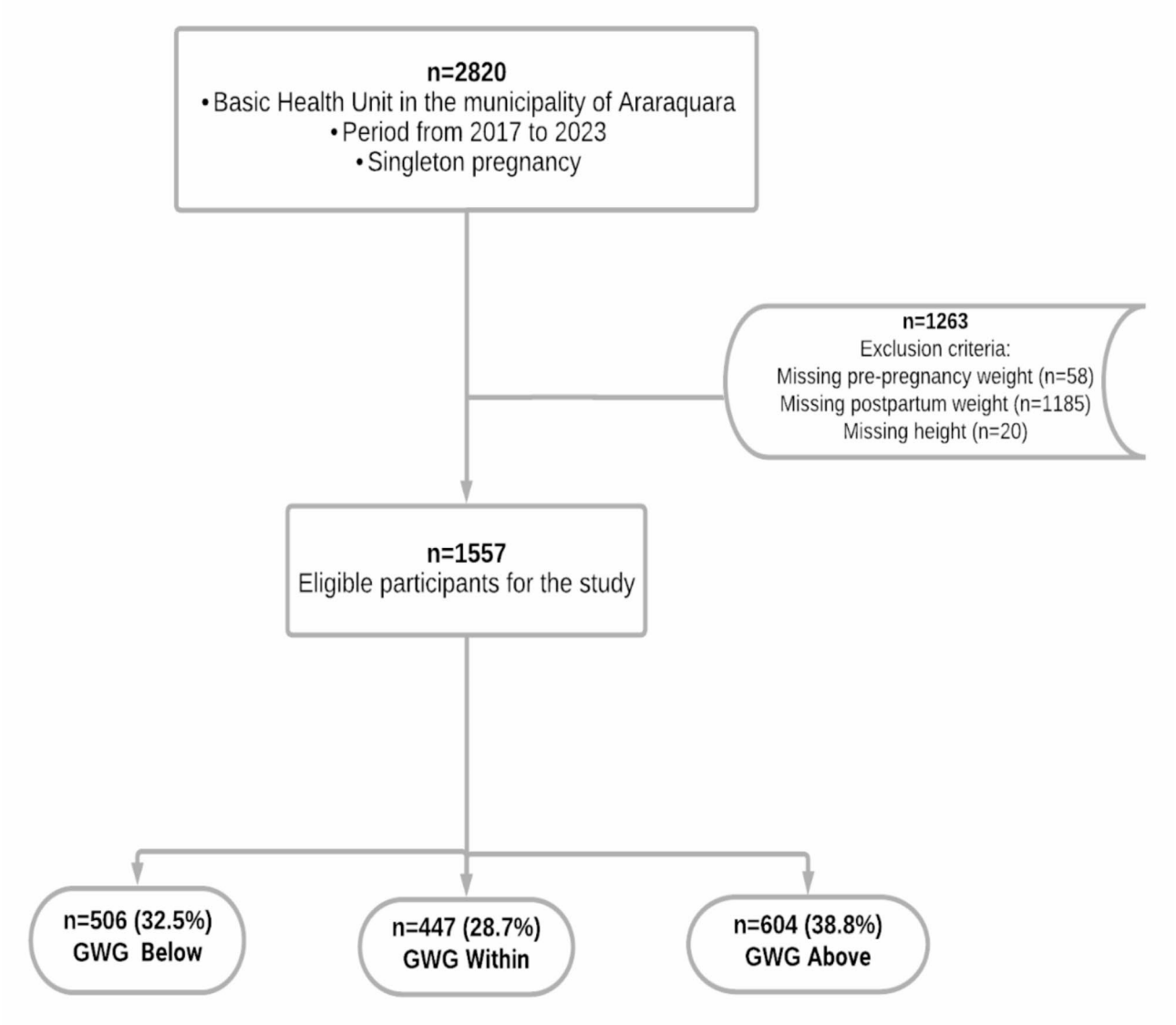

Fig. 2 Selection of the study population according to IOM recommendations

Table 1 Maternal characteristics associated with GWG according to IOM recommendations

Predictors	Overall 1557	Gestational Weight Gain (IOM-2009)			*P* value
		Within 447(28.7)	Below 506(32.5)	Above 604(38.7)	
Age (years) σ					
≤ 19	154(9.9)	47(3.02)	51(3.28)	56(3.6)	0.531
20–35	1189(76.4)	346(22.22)	389(25)	454(29.16)	
> 35	214(13.7)	54(3.47)	66(4.24)	94(6.04)	
Height(cm) σ					
1º tercil	534(34.34)	167(10.73)	187(12.03)	180(11.57)	0.003
2º tercil	505(32.48)	146(9.39)	170(10.93)	189(12.15)	
3º tercil	516(33.18)	134(8.62)	147(9.45)	235(15.11)	
Pre-gestational BMI (kg /m^2) τ	25.6(22.2–30.2)	25(21.3–28.6)	24.8(21.8–30.2)	26.8(23.2–31.2)	< 0.001
Arm circumference(cm) σ					
< 23	67(4.37)	23(1.50)	29(1.90	15(0.89)	< 0.001
23–28	474(31)	147(9.61)	190(12.42)	137(8.95)	
> 28	989(64.640	264(17.25)	283(18.50)	442(28.89)	
Body fat (%) τ	33.3(28.3–37.8)	32.3(26.9–36.6)	32.3(26.6–37)	34.7(30.3–39.1)	< 0.001
Gestational age (weeks) τ	39.4(38.5–40.3)	39.4(38.7–40.3)	39.2(38.1–40.1)	39.7(38.9–40.4)	
Maternal education (years) σ					
≤ 4	10(0.6)	1(0.06)	5(0.32)	4(0.26)	< 0.001
5–11	1181(75.9)	342(21.97)	389(24.98)	450(28.9)	
≥ 12	365(23.5)	104(6.68)	111(7.13)	150(9.63)	
Per capita income (R$) τ	666.7(400–1000)	665.9(400–970)	600(382.4–1000)	668(466.6–1000)	0.002
Race σ					
White	722(46.3)	208(13.36)	223(14.32)	291(18.69)	0.392
Non-white	835(53.6)	239(15.35)	283(18.18)	313(20.1)	
Marital status σ					
Married or in a stable relationship	1359(87.3)	388(24.93)	441(28.32)	530(34.04)	0.896
Single, separated, or widowed	198(12.7)	59(3.79)	65(4.17)	74(4.75)	
Physical activity σ					
Adequate	175(11.2)	50(3.21)	59(3.794)	66(4.24)	0.951
Inadequate	524(33.7)	156(10.02)	172(11.05)	196(12.59)	
Smoking σ					
No	1434(92.1)	409(26.27)	449(28.84)	576(36.99)	< 0.001
Yes	123(7.9)	38(2.44)	57(3.66)	28(1.8)	
Alcohol consumption σ					
No	1238(79.5)	353(22.67)	401(25.75)	482(30.96)	0.885
Yes	319(20.5)	94(6.04)	105(6.74)	120(7.71)	
Diabetes σ					
No	1479(95,0)	429(27.55)	459(29.48)	591(37.96)	< 0.001
Yes	78(5)	18(1.16)	47(3.02)	13(0.83)	
Hypertension σ					
No	1448(93)	420(26.97)	470(30.19)	558(35.84)	0.608
Yes	109(7)	27(1.73)	36(2.31)	46(2.95)	
hs-CRP (ng/mL) τ	5.9(3.1–11.7)	5.1(3–10)	6.1(3.2–11.9)	6.5(3.0-12.6)	0.137
HOMA (uUI/mL) τ	1.36(0.9–2.1)	1.4(0.9–2.1)	1.3(0.99–2.1)	1.42(1-2.2)	0.094
Hemoglobin (g/dL) τ	12.5(12-13.1)	12.6(11.9–13.1)	12.4(11.8–13)	12.6(121 – 13.2)	0.002
Glycated hemoglobin %τ	5.1(4.9–5.3)	5.1(4.9–5.3)	5.1(4.9–5.3)	5(4.8–5.3)	0.059
Cholesterol (mg/dL) τ	173(151–196)	172(152–196)	172(149–194)	174(152–198)	0.526
HDL-c (mg/dL) τ	56(48–64)	56(49–64)	55(47–62)	56(49–65)	0.012
LDL-c(mg/dL) τ	95(77–113)	94(79–111)	94(76–112)	96(77–115)	0.639
Triglycerides (mg/dL) τ	104(81–133)	104(80–134)	106(85–137)	100(80–129)	0.13

Data are presented as number (percentage) and median and interquartile range (percentile 25 - percentile 75)

Statistical differences among gestational weight gain groups were tested with: Kruskal-Wallis test for continuous predictors and χ2 test, Fisher's test for categorical predictors. Predictor Type: Continuous - τ and Categorical-σ

Table 2 Predictive performance on test data of the best algorithm for each outcome with hyperparameter tuning

Model	Class	Hyperparameter Tuning	AUC-ROC	Acuracy	Recall	Specificity	Precision	F1	MCC
LightGBM	GWG Within	{'num_leaves': 31, 'learning_rate': 0.1}	0.79	0.75	0.58	0.83	0.62	0.60	0.42
XGBoost	GWG Within	{'n_estimators': 200, 'max_depth': 3, 'learning_rate': 0.1}	0.79	0.74	0.59	0.82	0.60	0.60	0.41
Random Forest	GWG Within	{'n_estimators': 100, 'max_depth': 10}	0.77	0.76	0.58	0.82	0.60	0.59	0.41
CatBoost	GWG Within	{'learning_rate': 0.1, 'iterations': 100, 'depth': 6}	0.77	0.75	0.60	0.82	0.61	0.61	0.42
LightGBM	GWG Below	{'num_leaves': 31, 'learning_rate': 0.1}	0.76	0.68	0.74	0.64	0.57	0.64	0.37
XGBoost	GWG Below	{'n_estimators': 200, 'max_depth': 3, 'learning_rate': 0.1}	0.76	0.68	0.69	0.68	0.58	0.63	0.36
CatBoost	GWG Below	{'learning_rate': 0.1, 'iterations': 100, 'depth': 6}	0.75	0.68	0.69	0.67	0.58	0.63	0.36
Random Forest	GWG Below	{'n_estimators': 100, 'max_depth': 10}	0.73	0.68	0.77	0.59	0.55	0.64	0.35
AdaBoost	GWG Below	{'n_estimators': 200, 'learning_rate': 0.1}	0.71	0.66	0.77	0.60	0.55	0.64	0.36
AdaBoost	GWG Within	{'n_estimators': 200, 'learning_rate': 0.1}	0.71	0.72	0.55	0.80	0.57	0.56	0.35
CatBoost	GWG Above	{'learning_rate': 0.1, 'iterations': 100, 'depth': 6}	0.61	0.69	0.36	0.85	0.49	0.42	0.23
XGBoost	GWG Above	{'n_estimators': 200, 'max_depth': 3, 'learning_rate': 0.1}	0.65	0.65	0.36	0.84	0.47	0.40	0.21
LightGBM	GWG Above	{'num_leaves': 31, 'learning_rate': 0.1}	0.62	0.62	0.30	0.85	0.44	0.36	0.17
AdaBoost	GWG Above	{'n_estimators': 200, 'learning_rate': 0.1}	0.57	0.57	0.22	0.89	0.43	0.29	0.13
Random Forest	GWG Above	{'n_estimators': 100, 'max_depth': 10}	0.60	0.59	0.24	0.89	0.47	0.32	0.17

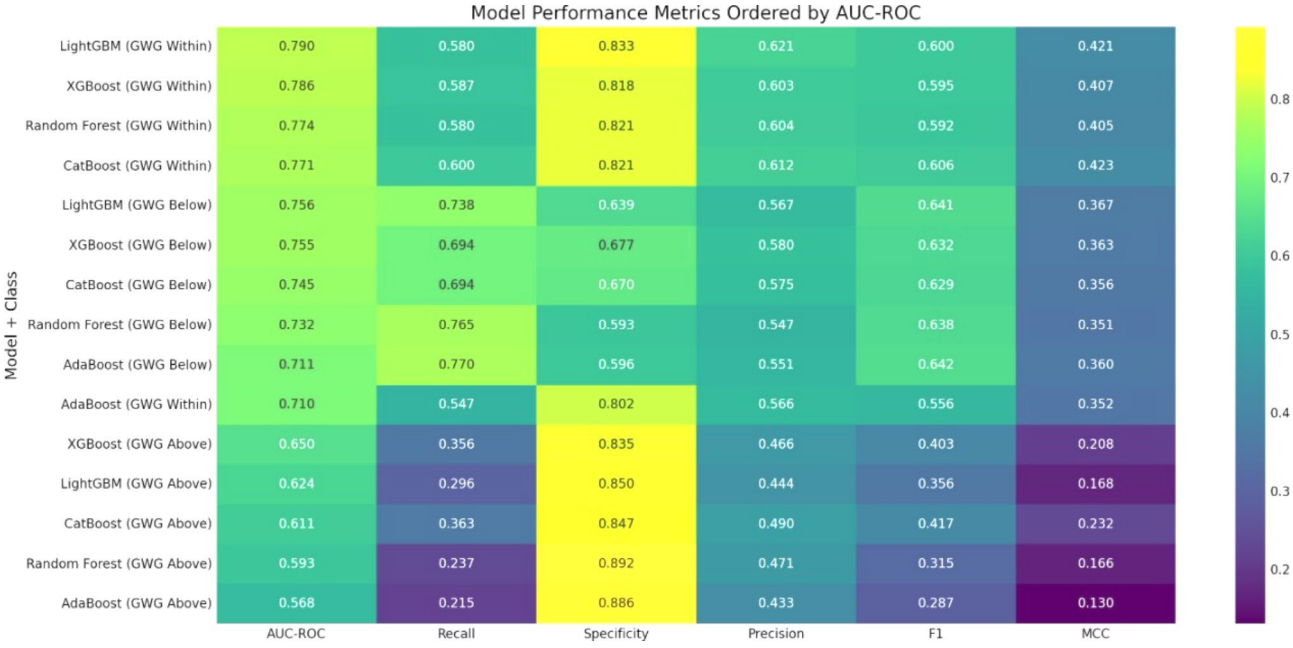

Fig. 3 Model performance metrics on test data ordered by AUC-ROC

Performance of predictive models

The XGBoost presented the best overall model, achieving an AUC-ROC of 0.79 for GWG within, 0.76 for GWG below, and 0.65 for GWG above. The LightGBM also performed well with an AUC-ROC of 0.79 for predicting GWG within recommendations, 0.76 for GWG below, and 0.624 for GWG above. Other algorithms, such as CatBoost, Random Forest, AdaBoost, and Logistic Regression, were also evaluated. (Table 2; Figs. 3, 4 and 5)

SHAP values and predictor importance

The use of SHAP values provided insight into the importance of various predictors for each GWG category. Pre-gestational BMI, maternal age, glycemic profile, hemoglobin levels, and arm circumference were identified as the most significant predictors. These variables were crucial in determining the likelihood of a pregnant woman falling into one of the GWG categories (below, within, or above IOM recommendations (Fig. 6).

Discussion

This study highlights the feasibility and utility of ML models in predicting GWG, thus providing a novel approach to enhancing prenatal care. The XGBoost and LightGBM models, in particular, exhibited strong predictive capabilities, with LightGBM achieving the highest AUC-ROC values across all GWG categories. These results align with existing literature that highlights the

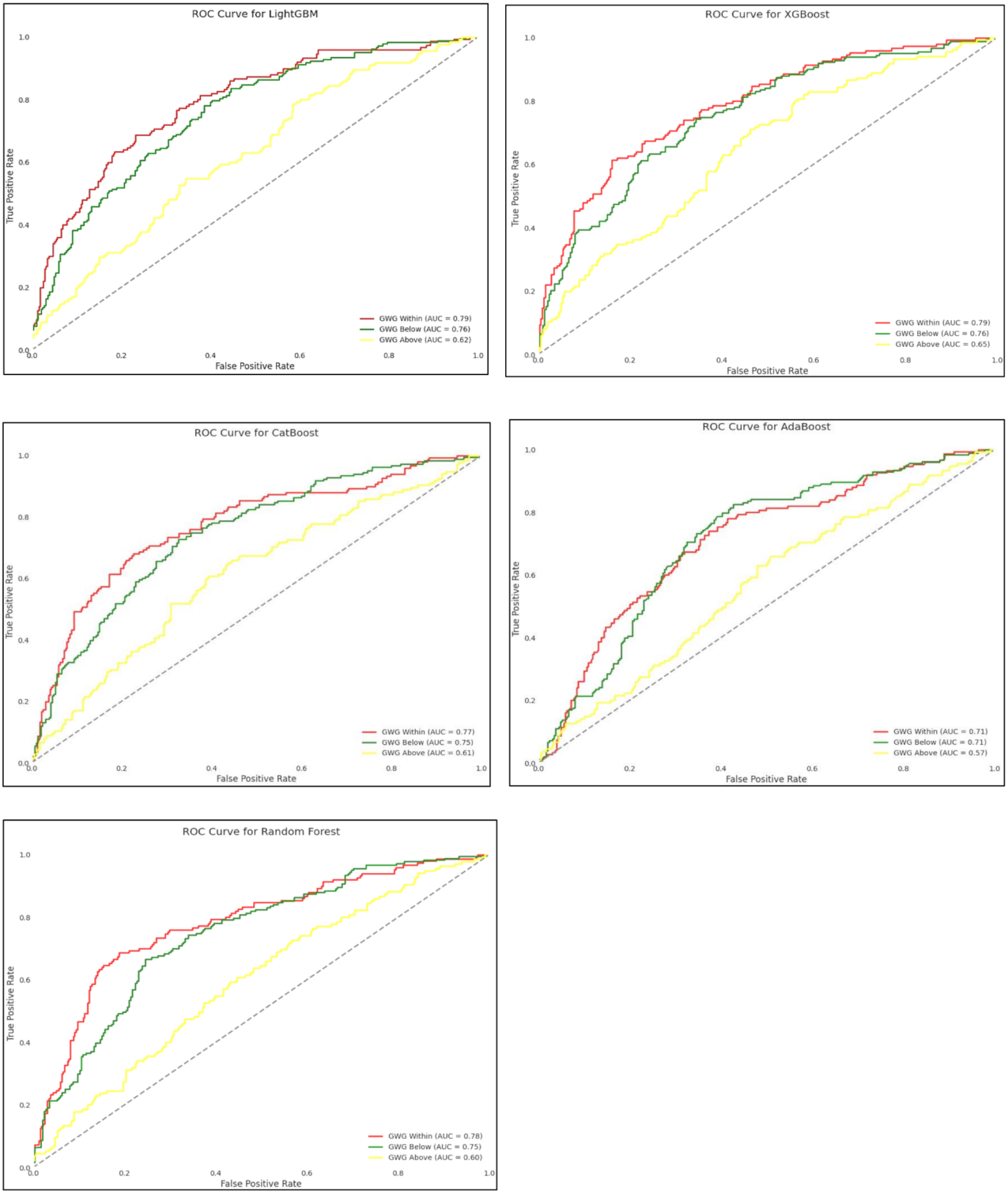

Fig. 4 Predictive performance on test data of algorithms for each model in Terms of Area Under the Receiver Operating Characteristic Curve (AUC-ROC)

superiority of gradient boosting algorithms for handling complex, non-linear relationships in large datasets . Recent studies indicate that boosting algorithms represent the state-of-the-art for tabular data demonstrating high performance across a wide range of tasks, including classification [35, 36].

The performance metrics, while robust, also indicate areas for improvement. For instance, the AUC-ROC

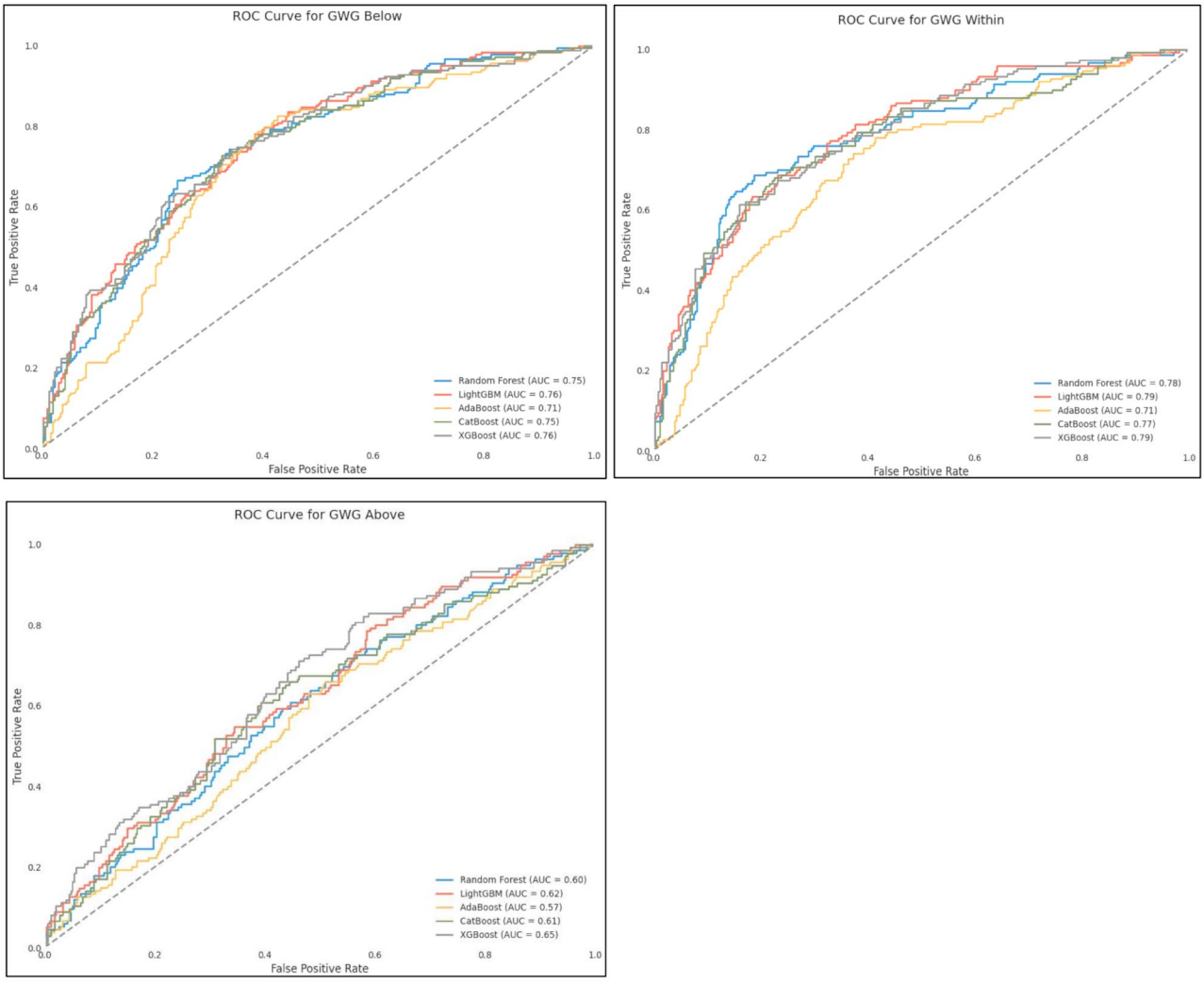

Fig. 5 ROC curve on test data for GWG prediction according to each class

values for predicting GWG above were lower compared to the other categories, suggesting a need for further refinement of the models to enhance their sensitivity to this particular outcome. Furthermore, integrating these ML models into clinical practice requires careful consideration of practical and ethical implications. Clinicians must be adequately trained to interpret and act on model predictions, and safeguards should be in place to ensure data privacy and security. The development of user-friendly interfaces and decision-support systems will be essential for the seamless integration of these tools into routine prenatal care.

This study highlights the feasibility and utility of ML models in predicting GWG, offering a valuable tool for early identification and management of at-risk pregnancies. By leveraging advanced analytics, healthcare providers can deliver more personalized and effective prenatal care, ultimately contributing to better health outcomes for mothers and their babies. Future research and clinical efforts should focus on refining these models, validating their applicability in diverse settings, and addressing the practical challenges associated with their implementation. The timely prediction and intervention, particularly starting in the second trimester, could significantly enhance pregnancy management and outcomes, supporting the findings of previous research on the importance of early GWG control [37, 38]. The ease of data collection for key predictors makes these models especially valuable for deployment in remote areas, broadening the impact and accessibility of advanced prenatal care solutions.

The significant predictors identified in this study, such as: pre-gestational BMI, maternal age, glycemic profile, hemoglobin levels, and arm circumference are consistent with known risk factors for GWG. These predictors collectively capture the multifaceted influences on GWG, encompassing physiological, demographic, and lifestyle factors. Importantly, these predictors are relatively easy to collect, even in remote or resource-limited settings,

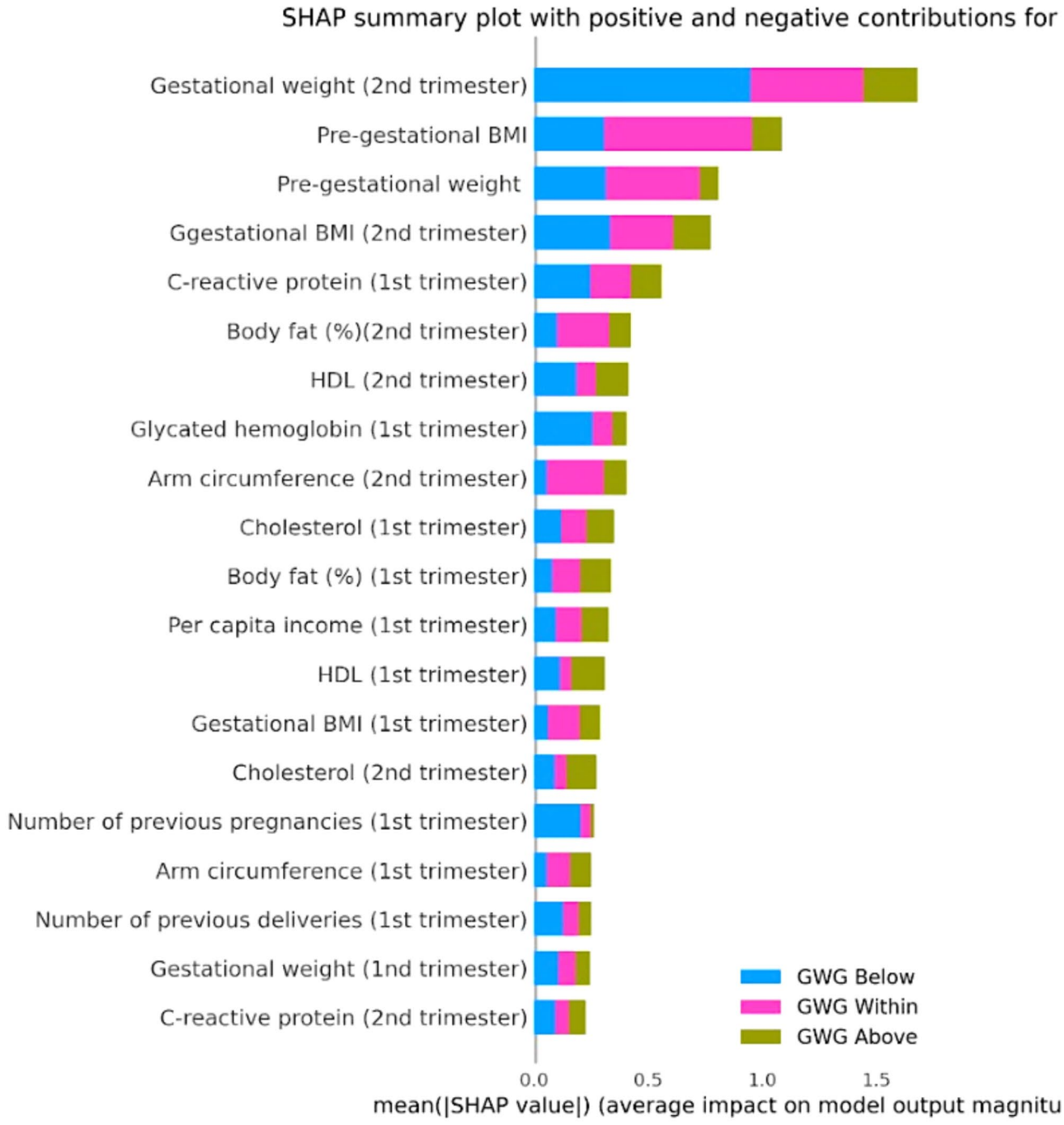

Fig. 6 Contributions of predictors to GWG SHAP summary plot for all classes

enhancing the feasibility of deploying these ML models in diverse clinical environments. The inclusion of these predictors enhances the model's ability to accurately stratify women based on their risk of inadequate or excessive GWG, thereby facilitating targeted interventions. Our results are consistent with other studies that have utilized ML to predict perinatal outcomes. For example, a study by Lee and Ahn (2020) demonstrated the effectiveness of models in predicting preterm birth, highlighting the importance of early and accurate predictions for timely intervention [12]. Similarly, Ramakrishnan, Rao, and He (2021) emphasized the potential of ML in identifying high-risk pregnancies and improving maternal-fetal health outcomes through early detection and personalized care [14].

The ease of collecting the significant predictors identified in this study makes these models particularly valuable for deployment in remote and resource-limited

areas. In such settings, where access to advanced medical infrastructure may be limited, the ability to gather basic anthropometric and clinical data can still enable effective risk stratification and intervention. Despite the promising results, several limitations must be acknowledged. The study's cohort is limited to a single geographic region (Araraquara, Brazil), which may affect the generalizability of the findings. Future research should aim to validate these models in diverse populations to ensure broader applicability. Additionally, the use of mean imputation to handle missing data, while standard in machine learning, may reduce the variability and precision of predictions. Future studies should explore more advanced imputation techniques to improve the robustness of the model.

Conclusion

This study demonstrates that machine learning models, particularly LightGBM and XGBoost, can effectively predict GWG categories. However, we acknowledge that the model's performance in predicting weight gain above the guidelines was limited and requires improvement. Integrating these models into clinical practice allows for the early identification of pregnant women at risk of inadequate or excessive weight gain, enabling timely and personalized interventions. While early detection can help reduce maternal and fetal complications, it is important to note that the applicability of these models depends on the availability of data, such as laboratory tests. Incorporating predictive models into clinical decision support systems can enhance prenatal care by offering more effective and individualized monitoring. However, in resource-limited settings where access to laboratory tests is restricted, the model's utility may be reduced. Future studies should consider adjusting and simplifying the variables to improve the model's applicability in different contexts, especially in low-resource regions.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12884-024-06952-8.

Supplementary Material 1

Acknowledgements

We especially thank the professionals, undergraduate, and graduate students who collaborated in the data collection for the Araraquara cohort. The authors would like to thank the São Paulo Research Foundation (FAPESP) for financial support (grant number 2015/03333-6) and also for the principal author's scholarship (grant number 2023/07936-3).

Author contributions

The authors' contributions were as follows AV: conceptualization, methodology. AV, FBF, GFS, AFC, LAL, and ADPC: investigation. AV, FF: Data analysis, AV, HG, FF, and GFS: visualization and writing - original draft. HG, ADPC, and PHCR: supervision, writing, review, and editing. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by the São Paulo Research Foundation (FAPESP) (grant number 2015/03333-6) AV has received a scholarship from São Paulo Research Foundation- FAPESP (grant number 2023/07936-3).

Data availability

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation. The code developed for constructing the algorithms along with the original dataset, is available on Github (https://github.com/Audency/Prediction-of-Gestational-Weight-Gain-for-Pregnancy.git).

Declarations

Ethics approval and consent to participate

The research received ethical approval from the Research Ethics Committee with Human Subjects at the School of Public Health, University of São Paulo, before the commencement of data collection, as per protocol CAEE: 59787216.2.0000.5421 and opinion number 1.885.874. All participants provided informed consent, consistent with the principles outlined in the Helsinki Declaration.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 28 May 2024 / Accepted: 4 November 2024
Published online: 08 November 2024

References

1. Victor A, de França da Silva Teles L, Aires IO, de Carvalho LF, Luzia LA, Artes R, et al. The impact of gestational weight gain on fetal and neonatal outcomes: the Araraquara Cohort Study. BMC Pregnancy Childbirth. 2024;24:320.
2. Goldstein RF, Abell SK, Ranasinha S, Misso ML, Boyle JA, Harrison CL, et al. Gestational weight gain across continents and ethnicity: systematic review and meta-analysis of maternal and infant outcomes in more than one million women. BMC Med. 2018;16:153.
3. Macdonald-Wallis C, Tilling K, Fraser A, Nelson SM, Lawlor DA. Gestational weight gain as a risk factor for hypertensive disorders of pregnancy. Am J Obstet Gynecol. 2013;209:e3271–32717.
4. Ren M, Li H, Cai W, Niu X, Ji W, Zhang Z, et al. Excessive gestational weight gain in accordance with the IOM criteria and the risk of hypertensive disorders of pregnancy: a meta-analysis. BMC Pregnancy Childbirth. 2018;18:281.
5. Truong YN, Yee LM, Caughey AB, Cheng YW. Weight gain in pregnancy: does the Institute of Medicine have it right ? The American Journal of Obstetrics & Gynecology. 2015;212:362.e1-362.e8.
6. Davis RR, Hofferth SL, Shenassa ED. Gestational weight gain and risk of infant death in the United States. Am J Public Health. 2014;104(Suppl 1):S90–5.
7. Voerman E, Santos S, Inskip H, Amiano P, Barros H, Charles MA, et al. Association of gestational weight gain with adverse maternal and infant outcomes. JAMA. 2019;321:1702–15.
8. Lipworth H, Barrett J, Murphy KE, Redelmeier D, Melamed N. Gestational weight gain in twin gestations and pregnancy outcomes: a systematic review and meta-analysis. BJOG. 2022;129:868–79.
9. Champion ML, Harper LM. Gestational weight gain: update on outcomes and interventions. Curr Diab Rep. 2020;20:11.
10. Gesche J, Nilas L. Pregnancy outcome according to pre-pregnancy body mass index and gestational weight gain. Int J Gynecol Obstet. 2015;129:240–3.
11. Jordan MI, Mitchell TM. Machine learning: trends, perspectives, and prospects. Science (1979). 2015;349:255–60.
12. Lee K-S, Ahn KH. Application of artificial intelligence in early diagnosis of spontaneous preterm labor and birth. Diagnostics (Basel). 2020;10.
13. Ayodele TO. Machine learning overview. New Adv Mach Learn. 2010;2:9–18.
14. Ramakrishnan R, Rao S, He J-R. Perinatal health predictors using artificial intelligence: a review. Womens Health (Lond). 2021;17:17455065211046132.

15. Batista AFM, Diniz CSG, Bonilha EA, Kawachi I, Chiavegatto Filho ADP. Neonatal mortality prediction with routinely collected data: a machine learning approach. BMC Pediatr. 2021;21:322.
16. Arayeshgari M, Najafi-Ghobadi S, Tarhsaz H, Parami S, Tapak L. Machine learning-based classifiers for the prediction of low birth weight. Healthc Inf Res. 2023;29:54–63.
17. Raj Pandey S, Ma J, Lai C-H, Raj Regmi P. A supervised machine learning approach to generate the auto rule for clinical decision support system. Trends Med. 2020;20.
18. Ranjbar A, Montazeri F, Farashah MV, Mehrnoush V, Darsareh F, Roozbeh N. Machine learning-based approach for predicting low birth weight. BMC Pregnancy Childbirth. 2023;23.
19. Naimi AI, Platt RW, Larkin JC. Machine learning for fetal growth prediction. Epidemiology. 2018;29:290–8.
20. Islam MN, Mustafina SN, Mahmud T, Khan NI. Machine learning to predict pregnancy outcomes: a systematic review, synthesizing framework and future research agenda. BMC Pregnancy Childbirth. 2022;22:348.
21. Martínez-Hortelano JA, Cavero-Redondo I, Álvarez-Bueno C, Garrido-Miguel M, Soriano-Cano A, Martínez-Vizcaíno V. Monitoring gestational weight gain and prepregnancy BMI using the 2009 IOM guidelines in the global population: a systematic review and meta-analysis. BMC Pregnancy Childbirth. 2020;20:649.
22. IOM. Weight gain during pregnancy: reexamining the guidelines. Washington, D.C.: National Academies; 2009.
23. Kang H. The prevention and handling of the missing data. Korean J Anesthesiology. 2013;64:402–6.
24. García-Laencina PJ, Sancho-Gómez JL, Figueiras-Vidal AR. Pattern classification with missing data: a review. Neural Comput Appl. 2010;19:263–82.
25. Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A. CatBoost: unbiased boosting with categorical features. Adv Neural Inf Process Syst. 2018;31.
26. Chen T, Guestrin C, Xgboost. A scalable tree boosting system. In: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. 2016. pp. 785–94.
27. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W et al. Lightgbm: a highly efficient gradient boosting decision tree. Adv Neural Inf Process Syst. 2017;30.
28. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. J Mach Learn Res. 2011;12:2825–30.
29. Bergstra J, Yamins D, Cox D. Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures. In: International conference on machine learning. PMLR; 2013. pp. 115–23.
30. Kursa MB, Rudnicki WR. Feature selection with the Boruta package. J Stat Softw. 2010;36:1–13.
31. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From local explanations to global understanding with explainable AI for trees. Nat Mach Intell. 2020;2:56–67.
32. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. Adv Neural Inf Process Syst. 2017;30.
33. Rodríguez-Pérez R, Bajorath J. Interpretation of machine learning models using shapley values: application to compound potency and multi-target activity predictions. J Comput Aided Mol Des. 2020;34:1013–26.
34. Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med. 2015;162:W1–73.
35. Shwartz-Ziv R, Armon A. Tabular data: deep learning is not all you need. Inform Fusion. 2022;81:84–90.
36. Borisov V, Leemann T, Seßler K, Haug J, Pawelczyk M, Kasneci G. Deep neural networks and tabular data: a survey. IEEE Trans Neural Netw Learn Syst. 2022:1–21.
37. Kominiarek MA, O'Dwyer LC, Simon MA, Plunkett BA. Targeting obstetric providers in interventions for obesity and gestational weight gain: a systematic review. PLoS ONE. 2018;13:e0205268.
38. Ren P, Yang XJ, Railton R, Jendza J, Anil L, Baidoo SK. Effects of different levels of feed intake during four short periods of gestation and housing systems on sows and litter performance. Anim Reprod Sci. 2018;188:21–34.

Publisher's note