

Interpretable Machine Learning Prediction Model for Prenatal Ambient Air Pollution Exposure and Its Impact on Small-for-Gestational-Age Births: The Korean CHildren's ENvironmental health Study (Ko-CHENS)

Payam Hosseinzadeh Kasani^{1,2}, Eun Sun Kim^{1,2}, Heui Seung Jo^{1,2}, <u>Kee Hyun Cho^{1,2}</u>, Woo Jin Kim³

¹ Department of Pediatrics, Kangwon National University Hospital, Chuncheon, Republic of Korea. ² Department of Pediatrics, Kangwon National University School of Medicine, Chuncheon, Republic of Korea.³ Department of Internal Medicine and Environmental Health Center, Kangwon National University, Chuncheon, Republic of Korea

Abstract

Introduction

Epidemiological studies have demonstrated an association between prenatal exposure to ambient air pollution and adverse birth outcomes. However, no predictive model assesses the risk of small-for-gestational-age (SGA) births in women exposed to specific pollutants during pregnancy, utilizing two pollution estimation methodologies: the Tele-Monitoring System (TMS) and the Kriging Interpolation Method (KRIG).

Results

Table 1. XGBoost model interpretation using feature importance based on SHAP ranking in in weight-based datasets.

Tele-Monitoring System (Weight-based Dataset)						K	Kriging Interpolation Method (Weight-based Dataset)					
1 st Trimester		2 nd Trimester		3 rd Trimester		1 st Trimester		2 nd Trimester		3 rd Trimester		
Feature	Importance	Feature	Importance	Feature	Importance	Feature	Importance	Feature	Importance	Feature	Importance	
TMS.NO2	2.7013	Pre_Preg Wt	4.7482	Maternal Ht	177.8706	GA	2.5448	KRIG.PM10	7.5009	GA	5.8274	
TMS.PM2.5	2.6251	TMS.NO2	4.6149	GA	64.9483	Pre_Preg Wt	2.3422	KRIG.NO2	6.9528	KRIG.PM2.5	5.2604	
GA	2.603	TMS.PM10	4.0815	TMS.PM10	14.9537	KRIG.CO.	2.2767	KRIG.03	6.1259	Pre_Preg Wt	5.0771	
Pre_Preg Wt	2.4704	TMS.PM2.5	4.0653	Mother Age	14.5105	KRIG.PM2.5	2.1299	KRIG.CO	5.4288	KRIG.PM10	5.0164	
TMS.SO2	2.4204	GA	4.0618	Pre_Preg Wt	12.9402	KRIG.NO2	2.0794	Pre_Preg Wt	5.3418	KRIG.CO	4.7744	
TMS.CO	2.0808	TMS.CO	3.7697	TMS.PM2.5	12.3359	KRIG.03	1.8684	GA	4.8263	Maternal Ht	3.7686	
TMS.PM10	1.9937	Maternal Ht	3.3628	TMS.O3	11.6533	KRIG.PM10	1.7798	KRIG.PM2.5	4.5774	Mother Age	3.1486	
TMS.O3	1.986	TMS.SO2	3.0578	TMS.CO	11.3532	Mother Age	1.5051	Maternal Ht	4.0647	KRIG.03	2.7708	
Maternal Ht	1.981	Mother Age	2.7385	TMS.NO2	7.1823	Maternal Ht	1.4588	C_section	3.6817	KRIG.NO2	2.1783	
Mother Age	1.5599	TMS.O3	2.6085	C_section	3.0948	Sex	1.4193	Mother Age	3.1899	Gravidity	1.6133	
Parity	1.4272	C_section	1.9641	Gravidity	2.9998	KRIG.SO2	1.2756	Gravidity	3.157	Sex	1.568	
C_section	1.1528	Sex	1.8353	Parity	2.7098	C_section	0.9625	KRIG.SO2	2.0644	C_section	1.5162	
Gravidity	0.9603	Parity	1.4014	TMS.SO2	2.0103	Parity	0.591	Sex	1.8445	KRIG.SO2	1.3131	
Sex	0.8933	Gravidity	1.0648	Sex	1.9663	Gravidity	0.5314	Parity	1.1604	Parity	1.2795	
GDM	0.074	PIH	0	GDM	0.045	PIH	0.0218	GDM	0.0374	PIH	0.0284	
PIH	0	GDM	0	PIH	0	GDM	0	PIH	0	GDM	0	

Methods:

Using data from 2,734 mothers enrolled in the Korean CHildren's ENvironmental Health Study (Ko-CHENS) cohort, we constructed two growth measure datasets (weightbased datasets (n = 2,734) 27.40% SGA, and length-based dataset (n = 2,422) with 16.52% SGA). We compared TMS and KRIG, applied by trimester and the entire pregnancy. Four machine learning models were evaluated by receiveroperating characteristic (ROC) curves, with Shapley Additive Explanations (SHAP) used to provide global model interpretations.

Results:

Our analysis showed no multicollinearity and XGBoost consistently outperformed other models across all trimesters and the entire pregnancy in both weight-based and length-based datasets, especially with the KRIG method. The highest performance was observed with KRIG for XGBoost in the first trimester for weight-based datasets (AUC 91.22%) and in the second trimester for length-based datasets (AUC 93.64%). The explainable results revealed by KRIG method the ambient pollution variables consistently appearing across all trimesters using the KRIG Method were particulate matter (PM2.5), nitrogen dioxide, and ozone in the weight-based dataset, and ozone, carbon monoxide (CO), and PM2.5 in the length-based dataset.

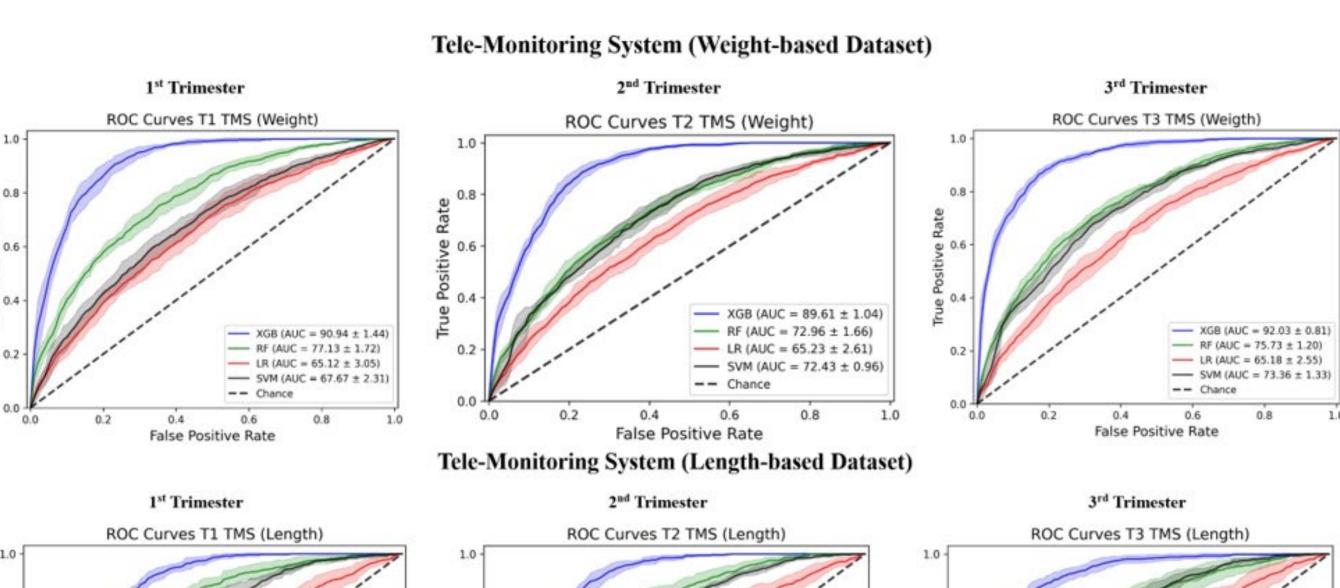
Table 2. XGBoost model interpretation using feature importance based on SHAP ranking in in length-based datasets.

Tele-Monitoring System (Length-based Dataset)						Kriging Interpolation Method (Length-based Dataset)					
1 st Trimester		2 nd Trimester		3 rd Trimester		1 st Trimester		2 nd Trimester		3 rd Trimester	
Feature	Importance	Feature	Importance	Feature	Importance	Feature	Importance	Feature	Importance	Feature	Importance
TMS.SO2	5.3315	TMS.NO2	6.6596	TMS.PM2.5	4.6561	KRIG.03	5.0141	KRIG.03	5.7116	KRIG.PM2.5	4.3573
TMS.PM2.5	4.2663	TMS.SO2	4.9789	TMS.PM10	3.8532	GA	3.7492	KRIG.NO2	4.6596	KRIG.CO	4.2971
TMS.NO2	4.0852	GA	4.9747	GA	3.6595	KRIG.CO	3.7144	KRIG.CO	4.6043	KRIG.PM10	3.8116
TMS.CO	3.7858	TMS.PM2.5	4.7738	TMS.CO	3.2641	KRIG.NO2	3.6967	KRIG.PM2.5	4.5273	GA	3.2887
TMS.PM10	3.2106	TMS.PM10	4.0984	Pre_Preg Wt	3.0887	KRIG.PM2.5	3.6714	GA	4.1688	Pre_Preg Wt	3.2256
Pre_Preg Wt	3.1363	TMS.CO	4.0809	Mother Age	2.7376	KRIG.PM10	3.3412	Pre_Preg Wt	3.5971	Mother Age	3.1029
Maternal Ht	3.009	Mother Age	3.6065	Maternal Ht	2.4393	KRIG.SO2	2.8925	KRIG.PM10	3.3025	Maternal Ht	2.6873
GA	2.6731	TMS.O3	3.561	TMS.O3	1.6235	Pre_Preg Wt	2.6476	Mother Age	2.8424	KRIG.03	2.0757
TMS.O3	2.5389	Pre_Preg Wt	3.4316	TMS.NO2	1.5398	Mother Age	2.6248	Maternal Ht	2.7613	Parity	1.8856
Mother Age	2.4838	Maternal Ht	3.0902	Parity	1.5099	Maternal Ht	2.5189	KRIG.SO2	2.0649	Gravidity	1.5811
Gravidity	1.6239	Gravidity	2.3839	C_section	1.4758	Gravidity	1.4176	Gravidity	1.4817	C_section	1.2522
C_section	1.4304	Sex	1.4252	Gravidity	1.2773	C_section	1.3574	C_section	1.3524	Sex	1.1702
Parity	1.1752	C_section	0.8561	Sex	0.8999	Sex	0.988	Sex	1.254	KRIG.NO2	1.1097
Sex	0.9212	Parity	0.6352	TMS.SO2	0.8626	Parity	0.7367	Parity	0.2736	KRIG.SO2	0.9832
PIH	0	GDM	0.0754	GDM	0.0434	GDM	0.0984	PIH	0	GDM	0.0157
GDM	0	PIH	0	PIH	0	PIH	0	GDM	0	PIH	0

Conclusion:

Machine learning algorithms can create effective tools for predicting SGA in mothers exposed to ambient air pollution, potentially aiding in identifying high-risk mothers and neonates.





— XGB (AUC = 85.87 ± 1.68) — XGB (AUC = 87.12 ± 1.47) — XGB (AUC = 88.68 ± 1.75 ----- RF (AUC = 78.83 ± 2.81) ----- RF (AUC = 78.49 ± 2.95) — RF (AUC = 74.98 ± 3.41) ----- LR (AUC = 61.51 ± 3.09) ---- LR (AUC = 61.28 ± 3.10) LR (AUC = 60.90 ± 4.53)

Wieght based datasets							Length based datasets						
1 st Trimester 2 ⁿ		2 nd Tr	2nd Trimester SGA (Weight)		3rd Trimester SGA (Weight)		1 st Trimester SGA (Length)		2nd Trimester SGA (Length)		3 rd Trimester		
	SGA (Weight)										SGA (Length)		
SGA (Weight)	1	SGA (Weight)	1	SGA (Weight)	1	SGA (Length)	1	SGA (Length)	1	SGA (Length)	1		
GA	0.089	GA	0.089	GA	0.089	KRIG_03	0.041	Sex	0.021	TMS.CO.T3	0.047		
Sex	0.054	Sex	0.054	Sex	0.054	TMS_PM10	0.022	GA	0.019	TMS.NO2.T3	0.039		
TMS_O3	0.032	TMS_O3	0.016	KRIG_CO	0.028	Sex	0.021	C_section	0.0089	KRIG.PM2.5.T3	0.034		
KRIG_03	0.022	- KRIG_SO2	0.0072	- KRIG_NO2	0.024	TMS_O3	0.02	- GDM	0.0089	KRIG.PM10.T3	0.032		
	0.011		0.0065		0.023	GA	0.019	KRIG_03	0.007	LUR.PM2.5.T3	0.031		
TMS_SO2		KRIG_03		TMS_CO						KRIG.CO.T3	0.029		
KRIG_NO2	0.0079	KRIG_NO2	0.0027	TMS_NO2	0.016	KRIG_PM10	0.016	TMS_03	-0.0022	LUR.PM10.T3	0.025		
PIH	0.0023	PIH	0.0023	TMS_SO2	0.0045	C_section	0.0089	Mother Age	-0.0036	TMS.PM2.5.T3	0.025		
TMS_NO2	0.0022	TMS_NO2	-0.0031	KRIG_PM2.5	0.0038	GDM	0.0089	PIH	-0.0057	TMS.PM10.T3	0.024		
KRIG_PM10	-0.0009	TMS_SO2	-0.0036	PIH	0.0023	Mother Age	-0.0036	KRIG_SO2	-0.015	KRIG.NO2.T3	0.021		
KRIG_SO2	-0.0038	GDM	-0.0098	TMS_PM2.5	0.0014	KRIG_PM2.5	-0.0049	TMS_NO2	-0.019	Sex	0.021		
TMS_PM10	-0.0085	TMS_PM10	-0.021	KRIG_PM10	-0.0023	PIH	-0.0057	KRIG_NO2	-0.028	GA	0.019		
GDM	-0.0098	KRIG_PM10	-0.022	TMS_PM10	-0.0044	TMS_PM2.5	-0.0089	KRIG_PM10	-0.029	LUR.NO2.T3	0.017		
KRIG_CO	-0.012	KRIG_CO	-0.022	_ GDM	-0.0098	TMS_NO2	-0.01	TMS_CO	-0.031	C_section GDM	0.0089		
KRIG_PM2.5	-0.013	TMS_CO	-0.028	KRIG_SO2	-0.011	KRIG_NO2	-0.025	TMS_PM10	-0.033	Mother Age	-0.0036		
TMS_CO	-0.017	KRIG_PM2.5	-0.032	TMS_03	-0.015	TMS_SO2	-0.025	TMS_PM2.5	-0.037	PIH	-0.0057		
TMS_PM2.5	-0.021	TMS_PM2.5	-0.035	KRIG_03	-0.024	TMS_CO	-0.033	KRIG_PM2.5	-0.041	TMS.SO2.T3	-0.014		
Mother Age	-0.04	Mother Age	-0.04	Mother Age	-0.04	KRIG_SO2	-0.045	TMS_SO2	-0.047	KRIG.SO2.T3	-0.018		
180.05 No. 19 K. MA				market and the		KRIG_CO	-0.059	KRIG_CO	-0.055	TMS.O3.T3	-0.034		
C_section	-0.054	C_section	-0.054	C_section	-0.054					KRIG.03.T3	-0.039		
Gravidity	-0.095	Gravidity	-0.095	Gravidity	-0.095	Pre_Preg Wt	-0.059	Pre_Preg Wt	-0.059	Pre_Preg Wt	-0.059		
Parity	-0.11	Parity	-0.11	Parity	-0.11	Maternal Ht	-0.064	Maternal Ht	-0.064	Maternal Ht	-0.064		
Maternal Ht	-0.11	Maternal Ht	-0.11	Maternal Ht	-0.11	Gravidity	-0.086	Gravidity	-0.086	Gravidity	-0.086		
Pre_Preg Wt	-0.16	Pre_Preg Wt	-0.16	Pre_Preg Wt	-0.16	Parity	-0.098	Parity	-0.098	Parity	-0.098		
	SGA (Weight)		SGA (Weight)		SGA (Weight)		SGA (Length)		SGA (Length)		SGA (Length)		

Figure 3. Feature correlation analysis. Correlation of variables with SGA outcome in 1st trimester; 2nd trimester; 3rd trimester. Positive impact sizes are represented by hues of red, while negative effect sizes are represented by shades of blue.

Discussion

- The consistent superior performance of the XGBoost model in our ROC curve analysis across all trimesters reveals its robustness and suggests that it can effectively manage the nonlinearities and interactions between various predictors in prenatal environments. - The variation in feature importance across different trimesters and datasets highlights the dynamic nature of factors influencing fetal growth. The fact that pollutants like PM2.5 and sulfur dioxide fluctuate in influence suggests that environmental risks may need trimester-specific public health interventions. This could guide more targeted prenatal care and policy adjustments depending on local environmental conditions. - The findings regarding gestational age reinforce its established importance in fetal monitoring but also suggest that integrating GA with real-time pollution data could refine risk assessments. Such integration could lead to the development of predictive models that are not only reactive but also proactive in adjusting to ongoing environmental and maternal health data throughout the pregnancy.

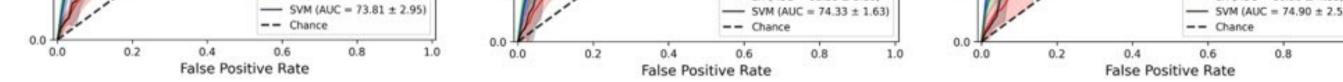


Figure 1. Receiver operating characteristic curves for 5-fold cross-validation for tele-monitoring system.

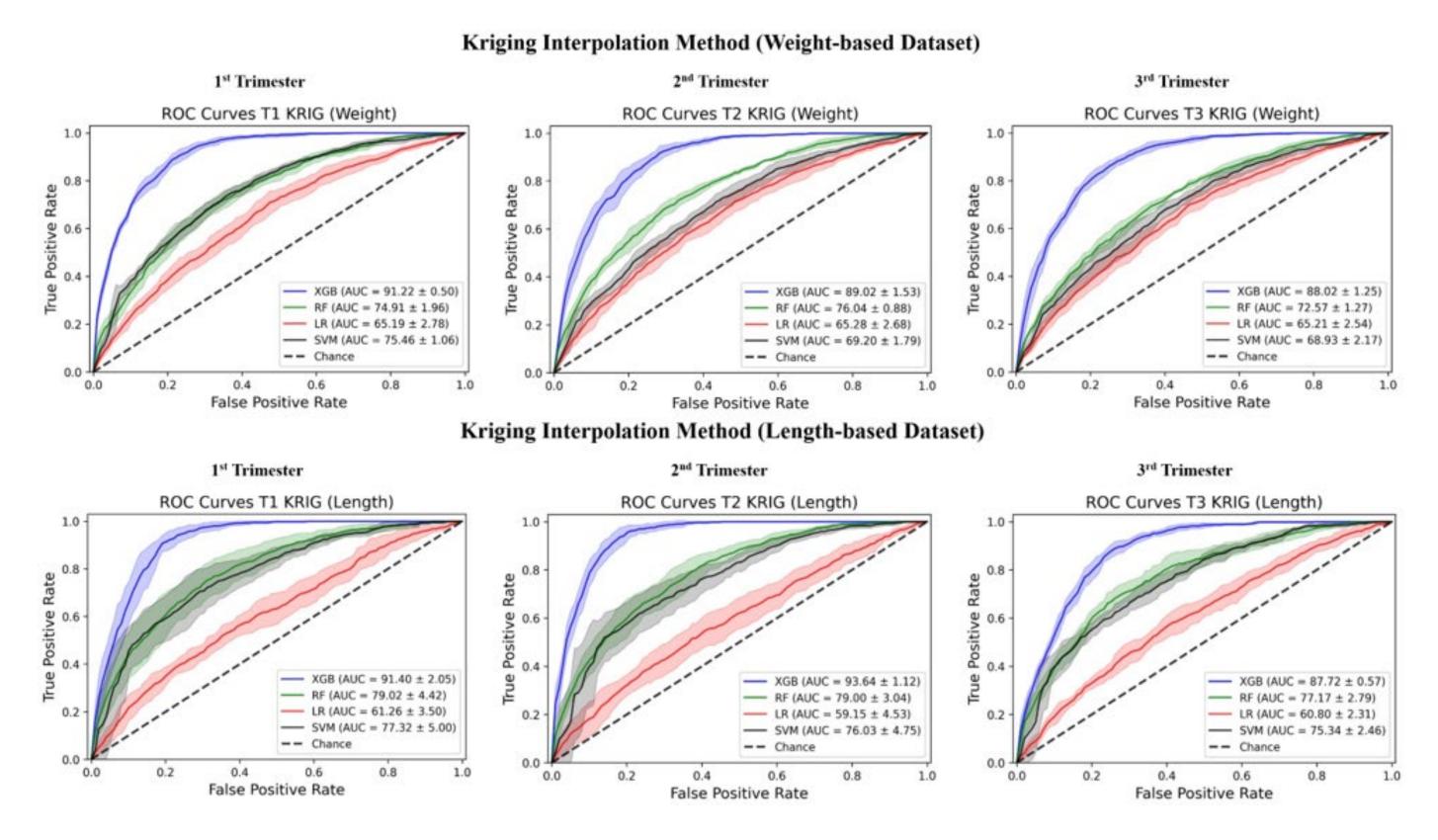


Figure 2. Receiver operating characteristic curves for 5-fold cross-validation for Kriging method.

References

Needleman HL, Schell A, Bellinger D, Leviton A, Allred EN. The Long-Term Effects of Exposure to Low Doses of Lead in Childhood. N Engl J Med [Internet]. 1990 Jan 11;322(2):83–8. Available from: http://www.nejm.org/doi/abs/10.1056/NEJM199001113220203 Pedersen M, Giorgis-Allemand L, Bernard C, Aguilera I, Andersen AMN, Ballester F, et al. Ambient air pollution and low birthweight: a European cohort study (ESCAPE). Lancet Respir Med [Internet]. 2013 Nov;1(9):695–704. Available from: https://linkinghub.elsevier.com/retrieve/pii/S2213260013701929

Jeong KS, Kim S, Kim WJ, Kim HC, Bae J, Hong YC, et al. Cohort profile: Beyond birth cohort study – The Korean CHildren's ENvironmental health Study (Ko-CHENS). Environ Res [Internet]. 2019 May;172:358–66. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0013935118306388

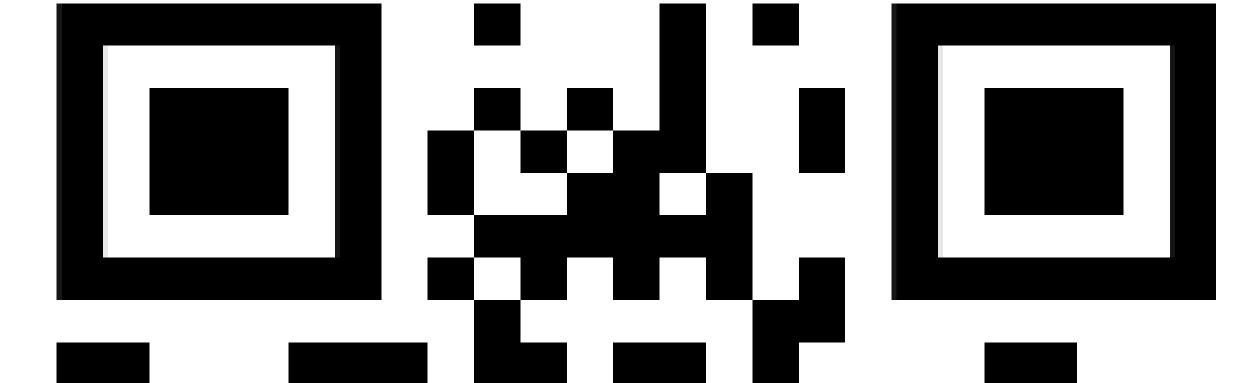




Interpretable Machine Learning Prediction Model for Prenatal Ambient Air Pollution Exposure and Its Impact on Small-for-Gestational-Age Births: The Korean CHildren's ENvironmental health Study (Ko-CHENS)

Video Presentation

Scan the QR code to watch the video presentation!



Hot Topics in Neonatology®