Machine learning-based risk factor analysis of adverse birth outcomes in very low birth weight infants

Hannah Cho1,2, Eun Hee Lee1, Kwang-Sig Lee3,4,5 & Ju Sun Heo1,2*

This study aimed to analyze major predictors of adverse birth outcomes in very low birth weight (VLBW) infants including particulate matter concentration (PM10), using machine learning and the national prospective cohort. Data consisted of 10,423 VLBW infants from the Korean Neonatal Network database during January 2013–December 2017. Five adverse birth outcomes were considered as the dependent variables, i.e., gestational age less than 28 weeks, gestational age less than 26 weeks, birth weight less than 1000 g, birth weight less than 750 g and small-for-gestational age. Thirty-three predictors were included and the artificial neural network, the decision tree, the logistic regression, the Naïve Bayes, the random forest and the support vector machine were used for predicting the dependent variables. Among the six prediction models, the random forest had the best performance (accuracy 0.79, area under the receiver-operating-characteristic curve 0.72).

According to the random forest variable importance, major predictors of adverse birth outcomes were maternal age (0.2131), birth-month (0.0767), PM10 month (0.0656), sex (0.0428), number of fetuses (0.0424), primipara (0.0395), maternal education (0.0352), pregnancy-induced hypertension (0.0347), chorioamnionitis (0.0336) and antenatal steroid (0.0318). In conclusion, adverse birth outcomes had strong associations with PM10 month as well as maternal and fetal factors.

Preterm birth brings a significant disease burden for infants and children in the world. On a yearly basis, fifteen million newborns are preterm on the globe and this is a major cause of one million deaths among those aged 0–4 years in the world1,2. The burden of preterm birth including its mortality and morbidity seems to be increasing in most countries3. Neonatal intensive care has advanced over the last few decades but extremely preterm (EP, less than 28 weeks of gestation) or extremely low birth weight (ELBW, less than 1000 g) infants have been subjected to death, neonatal complications or long-term neurodevelopmental impairment4–6, which presents a great challenge for the health care system as well7. Small for gestational age (SGA) is another significant contributor for the global burden of preterm birth8,9 and it is very important to identify the risk factors of these adverse birth outcomes above.

On the other hand, air pollution has become an important health issue in the world, increasing global mortality and morbidity. In 2016, ambient air pollution resulted in 4.2 million premature deaths worldwide10,11. This is particularly notable in a country such as Korea, where air pollution including particulate matter (PM) has registered rapid growth over the past few decades along with rapid industrialization and urbanization. Especially, maternal exposure to PM is reported to be a major cause of adverse health outcomes. Maternal exposure to PM1,2 (fine inhalable particles with diameters < 2.5 μm) or PM10 (inhalable particles with diameters < 10 μm) is found to have positive relationships with preterm birth12,13 and low birth weight14–16. However, little literature is available on associations between maternal exposure to PM and more adverse birth outcomes such as EP, ELBW or SGA. Furthermore, no endeavor has been made regarding the utilization of machine learning for the prediction of adverse birth outcomes among very low birth weight (VLBW, less than 1500 g) infants. In this context, this study employed machine learning and a national prospective cohort registry database to examine main predictors of adverse birth outcomes in VLBW infants including PM10 as a marker of air pollution.

1Department of Pediatrics, Korea University College of Medicine, Anam Hospital, 73 Goryeodae-ro, Seongbuk-gu, Seoul 02841, Korea. 2Department of Pediatrics, Korea University Anam Hospital, Seoul, Korea. 3AI Center, Korea University College of Medicine, Anam Hospital, 73 Goryeodae-ro, Seongbuk-gu, Seoul 02841, Korea. 4,5email: ecophy@hanmail.net; heojs08@korea.ac.kr
Results

Descriptive statistics are shown for participants' adverse birth outcomes and their predictors in Table 1. Among 10,423 participants, 3961 (38.0%), 1919 (18.4%), 3960 (38.0%), 1658 (15.9%) and 2242 (21.5%) belonged to the categories of GA < 28, GA = 26, BW < 1000, BW < 750 and SGA, respectively. Indeed, the mean and standard deviation of maternal age were 31 and 4.28, respectively. Monthly PM_{10} (micrograms per cubic meter) were 52 (January), 49 (February), 60 (March), 57 (April), 60 (May), 43 (June), 35 (July), 34 (August), 34 (September), 38 (October), 47 (November) and 48 (December). Yearly PM_{10} (micrograms per cubic meter), temperature average (degree Celsius), temperature min (degree Celsius) and temperature max (degree Celsius) in 2013, 2014, 2015, 2016 and 2017 were: 48, 46, 48, 48 and 47 for yearly PM_{10}; 12.85, 14.10, 13.85, 14.80 and 14.60 for yearly temperature average; 7.35, 8.55, 8.50, 10.05 and 9.30 for yearly temperature min; and 19.05, 20.50, 20.05, 20.45 and 20.65 for yearly temperature max. The results of univariate analysis were presented in Table 2. The P values were smaller than 0.10 for the following variables: birth-month with GA < 28, GA < 26, BW < 750 and SGA; sex (male) with GA < 28, GA < 26, BW < 1000 and BW < 750 (positive relationship); number of fetuses with GA < 28, GA < 26, BW < 1000 and BW < 750 (negative relationship); maternal age with SGA (positive relationship); and PM_{10} month with BW < 1000 and BW < 750 (positive relationship).

The performance of the random forest was the best among the six models in this study (Table 3). Its areas under the receiver-operating-characteristic curve, 0.73 (PM_{10} excluded) and 0.72 (PM_{10} included), were higher than its logistic-regression counterparts, 0.69 (PM_{10} excluded) and 0.68 (PM_{10} included). Based on random forest variable importance in Table 4 (PM_{10} excluded), main predictors of adverse birth outcomes were maternal age (0.2276), birth-year (0.1216), birth-month (0.1165), sex (0.0415), number of fetuses (0.0404), primipara (0.0373), maternal education (0.0361), pregnancy-induced hypertension (0.0325), choorioamnionitis (0.0319) and antenatal steroid (0.0307). Likewise, according to random forest variable importance in Table 5 (PM_{10} included), major predictors of adverse birth outcomes were maternal age (0.2131), birth-month (0.0767), PM_{10} month (0.0656), sex (0.0428), number of fetuses (0.0424), primipara (0.0395), maternal education (0.0352), pregnancy-induced hypertension (0.0347), choorioamnionitis (0.0336) and antenatal steroid (0.0319) (Fig. 1). These values were the pooled outcome for the five adverse birth outcomes. It was found that the variable importance of birth-year disappears after the inclusion of PM_{10} month in the model. Finally, Table 6 helps to understand how variable importance rankings vary among different adverse birth outcomes. The ranking of a top-5 (or top-10) predictor was highlighted with the color of orange (or mild blue) in each column of the five adverse birth outcomes in the table. Maternal age, birth-month and PM_{10} month were the first, second and third most important predictors across board. However, some predictors were outside the top 5 on average but within the top 5 in certain adverse birth outcomes: primipara (6th on average) ranked 4th and 5th in BW < 1000 and BW < 750, respectively; pregnancy-induced hypertension (8th on average) ranked 4th in SGA; and choorioamnionitis (9th on average) ranked 4th in GA < 28. Also, it needs to be noted that maternal education ranked within the top 10 across board.

Discussion

This study used machine learning to provide the most comprehensive investigation for the predictors of five adverse birth outcomes, using a national prospective cohort registry (KNN) database in VLBW infants. Among the six prediction models for adverse birth outcomes, the random forest had the best performance (accuracy 0.79, area under the receiver-operating-characteristic curve 0.72). According to the random forest variable importance, major predictors of adverse birth outcomes were maternal age, birth-month, PM_{10} month, sex, number of fetuses, primipara, maternal education, pregnancy-induced hypertension, choorioamnionitis and antenatal steroid.

In this study, birth month and PM_{10} had affect to EP, ELBW and SGA. Environmental changes according to the month such as air pollution, ambient temperature, the synthesis of vitamin D from sunlight could affect the adverse birth outcomes. Recently, existing literature reports the positive association of adverse birth outcomes with maternal exposure to PM during specific trimester or entire pregnancy. A meta-analysis study reported that PM_{10} was found to increase extremely preterm birth, and in particular, the incidence of preterm birth increased in Asia with high concentration of air pollution. Unlike previous studies, this study found that high air pollution exposure in the birth month, rather than specific trimester or entire pregnancy period, was strongly associated with adverse birth outcomes. This result was consistent with some previous studies that maternal exposure to PM during the birth month was a major risk factor for adverse birth outcomes. A time-series study conducted by Liu et al. showed that short term exposure to air pollution one week before childbirth increased preterm birth. In addition, Trasande et al.’s air pollution research model showed that PM_{10} concentration in the birth month was associated with low birth weight and VLBW. A possible pathway between maternal exposure to PM and adverse birth outcome would be oxidative stress, placental dysfunction, endothelial dysfunction and abnormal fetal growth. But more examination is to be done on this issue.

Sex (male) was the fourth most important predictor for adverse birth outcomes in VLBW infants in this study. This result is consistent with a recent review, which suggests the following mechanism: In the presence of a male fetus the trophoblast causes a more pro-inflammatory environment, and the maturation of the fetal hypophalamic–pituitary–adrenal axis and the expression of placental genes follow sex-dependent patterns. The number of fetuses was found to have a negative association with adverse birth outcomes in VLBW infants in this study. This finding does not agree with a recent review, which reports that the number of fetuses is a strong risk factor for adverse birth outcomes. One way to explain this discrepancy is that more active prenatal screening and monitoring were done in the case of multiple fetuses in this study: Pregnancy with multiple fetuses often involves in vitro fertilization, which can associate with high socioeconomic status and active prenatal care.

Maternal age was the first most important predictor for adverse birth outcomes in VLBW infants but the result of univariate analysis was significant only for SGA in this study. The participants of this study were limited to VLBW infants and more investigation based on broader participants is to be done for more conclusive result.
| Variable                                      | n   | %    |
|----------------------------------------------|-----|------|
| Gestational age < 28 weeks                   | 3961| 38.0 |
| Gestational age < 26 weeks                   | 1919| 18.4 |
| Birth weight < 1000 g                        | 3960| 38.0 |
| Birth weight < 750 g                         | 1658| 15.9 |
| Small-for-gestational-age                    | 2242| 21.5 |
| Sex: Male                                    | 5270| 50.6 |
| Birth-year                                   |     |      |
| 2013                                         | 1395| 13.4 |
| 2014                                         | 2126| 20.4 |
| 2015                                         | 2399| 23.0 |
| 2016                                         | 2365| 22.7 |
| 2017                                         | 2138| 20.5 |
| Birth-Season: Spring                         | 2535| 24.3 |
| Birth-Season: Summer                         | 2623| 25.2 |
| Birth-Season: Autumn                         | 2759| 26.5 |
| Birth-Season: Winter                         | 2506| 24.0 |
| Number of fetuses                            |     |      |
| 1                                            | 6761| 64.9 |
| 2                                            | 3247| 31.2 |
| 3                                            | 400 | 3.8  |
| 4 or more                                    | 15  | 0.1  |
| In vitro fertilization                       |     |      |
| Gestational DM                               | 830 | 8.0  |
| Overt DM                                     | 114 | 1.1  |
| Pregnancy-induced hypertension               | 1586| 19.2 |
| Chronic hypertension                         | 223 | 2.1  |
| Chorioamnionitis                             | 3019| 29.0 |
| PROM                                         | 3661| 35.1 |
| PROM > 18 h                                  | 2467| 23.7 |
| Antenatal steroid                            | 8310| 79.7 |
| Cesarean section                             | 8106| 77.8 |
| Oligohydramnios                              | 1413| 13.6 |
| Polyhydramnios                               | 153 | 1.5  |
| Primipara                                    | 6497| 62.3 |
| Maternal education                           |     |      |
| Elementary                                   | 29  | 0.3  |
| Junior high                                  | 122 | 1.2  |
| Senior high                                  | 1931| 18.5 |
| College or higher                            | 8341| 80.0 |
| Maternal citizenship                         |     |      |
| Korea                                        | 10,023| 96.2|
| China                                        | 124 | 1.2  |
| Vietnam                                      | 122 | 1.2  |
| Philippines                                  | 62  | 0.6  |
| Cambodia                                     | 35  | 0.3  |
| Other                                        | 57  | 0.2  |
| Paternal education                           |     |      |
| Elementary                                   | 8   | 0.1  |
| Junior high                                  | 52  | 0.5  |
| Senior high                                  | 1280| 12.3 |
| College or higher                            | 9083| 87.1 |
| Paternal citizenship                         |     |      |
| Korea                                        | 10,191| 97.8|
| Vietnam                                      | 79  | 0.8  |
| Philippines                                  | 19  | 0.2  |
| Cambodia                                     | 7   | 0.1  |
| Continued                                    |     |      |
Table 1. Descriptive statistics: adverse birth outcomes and categorical predictors. DM, diabetes mellitus; PROM, prelabor rupture of membrane.

| Variable                        | n   | %  |
|---------------------------------|-----|----|
| China                           | 5   | 0.0|
| Other                           | 122 | 1.2|
| Unmarried                       | 216 | 2.1|
| Congenital infection            | 128 | 1.2|

Table 2. Univariate analysis. BW, birth weight (grams); GA, gestational age (weeks); N-fetuses, number of fetuses; PM, particulate matter; SGA, small-for-gestational age. *P < 0.10 chi-square test for the equality of proportions “Yes” or T test for the equality of means.

| Variable       | GA < 28 | GA < 26 | BW < 1000 | BW < 750 | SGA      |
|----------------|---------|---------|-----------|----------|----------|
| Birth-Month (%)|         |         |           |          |          |
| January        | 0.065*  | 0.069*  | 0.115     | 0.002*   | 0.021*   |
| February       |         |         |           |          |          |
| March          |         |         |           |          |          |
| April          |         |         |           |          |          |
| May            |         |         |           |          |          |
| June           |         |         |           |          |          |
| July           |         |         |           |          |          |
| August         |         |         |           |          |          |
| September      |         |         |           |          |          |
| October        |         |         |           |          |          |
| November       |         |         |           |          |          |
| December       |         |         |           |          |          |
| Male (%)       | <0.001* | 0.051*  | 0.004*    | 0.001*   | 0.092    |
| N-Fetuses (%)  | <0.001* | 0.023*  | <0.001*   | 0.078*   | 0.704    |
| 1              | 62.4    | 68.9    | 62.9      | 64.5     | 64.6     |
| 2              | 32.8    | 28.4    | 32.4      | 31.4     | 31.4     |
| 3              | 4.5     | 2.7     | 4.4       | 3.9      | 3.9      |
| 4 or more      | 0.2     | 0.0     | 0.2       | 0.2      | 0.1      |
| Maternal age (mean ± SD)        | 32.99 ± 4.28 | 33.05 ± 4.27 | 33.02 ± 4.28 | 32.99 ± 4.29 | 0.453  |
| PM10 Month (mean ± SD)          | 46.12 ± 9.28 | 46.33 ± 9.34 | 46.44 ± 9.50 | 46.07 ± 9.24 | 0.083* |

Table 3. Model Performance. ANN artificial neural network; AUC, area under the receiver-operating-characteristic curve; BW, birth weight (grams); DT, decision tree; GA, gestational age (weeks); LR, logistic regression; NB, naive bayes; PM, particulate matter; RF, random forest—1000 trees; SGA, small-for-gestational age; SVM, support vector machine.
| Variable                          | $GA<28$ | $GA<26$ | $BW<1000$ | $BW<750$ | $SGA$ | Average | Rank |
|----------------------------------|---------|---------|-----------|----------|-------|---------|------|
| Sex (Male)                       | 0.0388* | 0.0460* | 0.0405*   | 0.0398*  | 0.0418 | 0.0415  | 4    |
| Birth-Year                       | 0.1207* | 0.1159  | 0.1286    | 0.1285   | 0.1141 | 0.1216  | 2    |
| Birth-Month                      | 0.1183* | 0.1166* | 0.1217    | 0.1163*  | 0.1098*| 0.1165  | 3    |
| Birth-Season: Spring             | 0.0152  | 0.0143  | 0.0151    | 0.0127   | 0.0141 | 0.0143  | 21   |
| Birth-Season: Summer             | 0.0162  | 0.0164  | 0.0165    | 0.0154   | 0.0150 | 0.0159  | 17   |
| Birth-Season: Autumn             | 0.0155  | 0.0150  | 0.0157    | 0.0144   | 0.0150 | 0.0151  | 19   |
| Birth-Season: Winter             | 0.0154  | 0.0141  | 0.0154    | 0.0129   | 0.0149 | 0.0146  | 20   |
| Number of fetuses                | 0.0402* | 0.0429* | 0.0345*   | 0.0435*  | 0.0410 | 0.0404  | 5    |
| In vitro fertilization           | 0.0282  | 0.0304  | 0.0284    | 0.0309   | 0.0249 | 0.0286  | 12   |
| Gestational DM                   | 0.0185  | 0.0131  | 0.0151    | 0.0117   | 0.0173 | 0.0152  | 18   |
| Overt DM                         | 0.0046  | 0.0038  | 0.0042    | 0.0031   | 0.0043 | 0.0040  | 28   |
| PIH                              | 0.0318  | 0.0204  | 0.0237    | 0.0283   | 0.0581 | 0.0325  | 8    |
| Chronic hypertension             | 0.0072  | 0.0067  | 0.0074    | 0.0095   | 0.0079 | 0.0077  | 23   |
| Chorioamnionitis                 | 0.0425  | 0.0402  | 0.0248    | 0.0235   | 0.0285 | 0.0319  | 9    |
| PROM                             | 0.0226  | 0.0228  | 0.0270    | 0.0287   | 0.0407 | 0.0284  | 13   |
| PROM > 18 Hour                   | 0.0191  | 0.0205  | 0.0207    | 0.0241   | 0.0237 | 0.0216  | 16   |
| Antenatal steroid                | 0.0297  | 0.0342  | 0.0268    | 0.0333   | 0.0297 | 0.0307  | 10   |
| Cesarean section                 | 0.0300  | 0.0328  | 0.0305    | 0.0240   | 0.0267 | 0.0288  | 11   |
| Oligohydramnios                  | 0.0254  | 0.0264  | 0.0193    | 0.0215   | 0.0453 | 0.0276  | 14   |
| Polyhydramnios                   | 0.0065  | 0.0056  | 0.0059    | 0.0054   | 0.0053 | 0.0057  | 26   |
| Maternal age                     | 0.2242  | 0.2268  | 0.2381    | 0.2363   | 0.2125*| 0.2276  | 1    |
| Primipara                        | 0.0351  | 0.0388  | 0.0434    | 0.0405   | 0.0287 | 0.0373  | 6    |
| Maternal education               | 0.0357  | 0.0383  | 0.0377    | 0.0364   | 0.0325 | 0.0361  | 7    |
| Maternal citizenship             | 0.0132  | 0.0100  | 0.0130    | 0.0120   | 0.0109 | 0.0118  | 22   |
| Paternal education               | 0.0251  | 0.0256  | 0.0262    | 0.0269   | 0.0244 | 0.0256  | 15   |
| Paternal citizenship             | 0.0079  | 0.0076  | 0.0073    | 0.0087   | 0.0065 | 0.0076  | 24   |
| Unmarried                        | 0.0067  | 0.0086  | 0.0069    | 0.0069   | 0.0041 | 0.0066  | 25   |
| Congenital infection             | 0.0055  | 0.0055  | 0.0056    | 0.0048   | 0.0024 | 0.0048  | 27   |

Table 4. Random forest variable importance for adverse birth outcomes: PM$_{10}$ excluded. The ranking of a top-5 predictor was highlighted with the color of orange and 6–10 predictor was highlighted with the color of mild blue. BW, birth weight (grams); DM, diabetes mellitus; GA, gestational age (weeks); PIH, pregnancy-induced hypertension; PM, particulate matter; PROM, prelabor rupture of membranes; SGA, small-for-gestational age. *$P<0.10$ chi-square or T Test (Table 2).

Some predictors were outside the top 5 on average but within the top 5 in certain adverse birth outcomes in this study: pregnancy-induced hypertension (eighth on average) ranked fourth in SGA; and chorioamnionitis (ninth on average) ranked fourth in $GA<28$. Chorioamnionitis is considered to be a major predictor of preterm birth.$^{30,31}$ It is expected to encourage pro-inflammatory cytokines, uterine contraction and preterm labor; or it is supposed to promote matrix metalloproteinase activation, fetal membrane degradation and preterm birth.$^{32}$ Also, early onset preeclampsia is reported to cause abnormal placentation and spiral artery remodeling, which in turn lead to restriction on blood flow to uterine arteries and abnormal fetal growth.$^{33–35}$

**Limitations.** Firstly, this study did not analyze possible mediating effects among predictors. Secondly, this study adopted the binary categories of adverse birth outcomes (no, yes), which can be extended to multiple categories with more clinical insights. Thirdly, it was beyond the scope of this study to investigate a variety of mechanisms between PM and adverse birth outcomes. Little effort has been made and more examination is needed in this direction. Fourthly, synthesizing different modes of machine learning methods for different types of adverse birth outcome data would break new ground on this topic. Fifthly, this study did not consider indoor environmental factors, which were reported as major predictors of adverse birth outcomes along with PM.$^{36,37}$

**Conclusions.** The current study is the first to evaluate the predictors including PM$_{10}$ month on adverse birth outcomes in VLBW infants using machine learning. Adverse birth outcomes such as EP, ELBW or SGA have strong associations with PM$_{10}$ month as well as maternal and fetal predictors among the VLBW infants. For the prevention of adverse effects on birth outcomes, clinical and/or policy measures are needed regarding these predictors.

**Methods**  
**Participants and variables.** Data consisted of 10,423 VLBW infants from the Korean Neonatal Network (KNN) database during January 2013–December 2017. The KNN started on April 2013 as a national prospective cohort registry of VLBW infants admitted or transferred to neonatal intensive care units across South Korea (It covers 74 neonatal intensive care units now). It collects the perinatal and neonatal data of VLBW infants based on a standardized operating procedure.$^{37}$

---

https://doi.org/10.1038/s41598-022-16234-y
| Variable                                      | GA<28 | GA<26 | BW<1000 | BW<750 | SGA    | Average | Rank |
|----------------------------------------------|-------|-------|---------|--------|--------|---------|------|
| Sex (Male)                                   | 0.0415* | 0.0463* | 0.0433* | 0.0373* | 0.0455 | 0.0428 | 4    |
| Birth-Year                                   | 0.0238 | 0.0246 | 0.0246  | 0.0257  | 0.0240 | 0.0245 | 16   |
| Birth-Month                                  | 0.0778* | 0.0754* | 0.0808  | 0.0783* | 0.0713* | 0.0767 | 2    |
| Birth-Season: Spring                         | 0.0105 | 0.0091 | 0.0102  | 0.0097  | 0.0096 | 0.0098 | 27   |
| Birth-Season: Summer                         | 0.0132 | 0.0129 | 0.0141  | 0.0139  | 0.0126 | 0.0133 | 23   |
| Birth-Season: Autumn                         | 0.0132 | 0.0131 | 0.0142  | 0.0129  | 0.0130 | 0.0133 | 24   |
| Birth-Season: Winter                         | 0.0126 | 0.0110 | 0.0122  | 0.0111  | 0.0116 | 0.0117 | 25   |
| Number of fetuses                            | 0.0431* | 0.0464* | 0.0349* | 0.0436* | 0.0443 | 0.0424 | 5    |
| In vitro fertilization                       | 0.0298 | 0.0330 | 0.0302  | 0.0319  | 0.0255 | 0.0301 | 12   |
| Gestational DM                               | 0.0169 | 0.0136 | 0.0150  | 0.0117  | 0.0152 | 0.0145 | 22   |
| Overt DM                                     | 0.0037 | 0.0039 | 0.0041  | 0.0026  | 0.0043 | 0.0037 | 33   |
| PIH                                          | 0.0354 | 0.0214 | 0.0274  | 0.0303  | 0.0592 | 0.0347 | 8    |
| Chronic hypertension                         | 0.0068 | 0.0064 | 0.0074  | 0.0091  | 0.0077 | 0.0075 | 28   |
| Chorioamnionitis                             | 0.0474 | 0.0390 | 0.0269  | 0.0255  | 0.0290 | 0.0336 | 9    |
| PROM                                         | 0.0248 | 0.0243 | 0.0277  | 0.0295  | 0.0436 | 0.0300 | 13   |
| PROM > 18 Hour                               | 0.0199 | 0.0217 | 0.0210  | 0.0244  | 0.0239 | 0.0222 | 20   |
| Antenatal steroid                             | 0.0299 | 0.0362 | 0.0263  | 0.0342  | 0.0323 | 0.0318 | 10   |
| Cesarean section                             | 0.0327 | 0.0325 | 0.0311  | 0.0273  | 0.0270 | 0.0301 | 11   |
| Oligohydramnios                              | 0.0257 | 0.0251 | 0.0219  | 0.0231  | 0.0434 | 0.0278 | 14   |
| Polysydramnios                               | 0.0056 | 0.0060 | 0.0056  | 0.0055  | 0.0055 | 0.0056 | 31   |
| Maternal age                                 | 0.2071 | 0.2174 | 0.2247  | 0.2193  | 0.1969* | 0.2131 | 1    |
| Primipara                                    | 0.0372 | 0.0395 | 0.0468  | 0.0426  | 0.0312 | 0.0395 | 6    |
| Maternal education                           | 0.0352 | 0.0363 | 0.0372  | 0.0359  | 0.0313 | 0.0352 | 7    |
| Maternal citizenship                         | 0.0127 | 0.0087 | 0.0122  | 0.0121  | 0.0107 | 0.0113 | 26   |
| Paternal education                           | 0.0250 | 0.0257 | 0.0257  | 0.0259  | 0.0238 | 0.0252 | 15   |
| Paternal citizenship                         | 0.0069 | 0.0062 | 0.0072  | 0.0085  | 0.0064 | 0.0070 | 29   |
| Unmarried                                    | 0.0061 | 0.0078 | 0.0061  | 0.0071  | 0.0045 | 0.0063 | 30   |
| Congenital infection                         | 0.0043 | 0.0050 | 0.0051  | 0.0044  | 0.0022 | 0.0042 | 32   |
| PM10 Year                                    | 0.0141 | 0.0148 | 0.0147  | 0.0156  | 0.0143 | 0.0147 | 21   |
| PM10 Month                                   | 0.0673 | 0.0651 | 0.0699* | 0.0651* | 0.0608 | 0.0656 | 3    |
| Temperature Year Average                     | 0.0231 | 0.0236 | 0.0239  | 0.0252  | 0.0230 | 0.0238 | 19   |
| Temperature Year Min                         | 0.0231 | 0.0236 | 0.0240  | 0.0254  | 0.0231 | 0.0238 | 18   |
| Temperature Year Max                         | 0.0236 | 0.0241 | 0.0241  | 0.0254  | 0.0234 | 0.0241 | 17   |

Table 5. Random forest variable importance for adverse birth outcomes: PM10 included. The ranking of a top-5 predictor was highlighted with the color of orange and 6–10 predictor was highlighted with the color of mild blue. BW, birth weight (grams); DM, diabetes mellitus; GA, gestational age (weeks); PIH, pregnancy-induced hypertension; PM, particulate matter; PROM, prelabor rupture of membranes; SGA, small-for-gestational age. *P<0.10 chi-square or T test (Table 2).

Five adverse birth outcomes were considered as binary dependent variables (no, yes), i.e., gestational age less than 28 weeks (GA<28), GA less than 26 weeks (GA<26), birth weight less than 1000 g (BW<1000), BW less than 750 g (BW<750) and SGA. Thirty-three predictors were included: sex—male (no, yes), birth-year (2013, 2014, 2015, 2016, 2017), birth-month (1, 2, …, 12), birth-season-spring (no, yes), birth-season-summer (no, yes), birth-season-autumn (no, yes), birth-season-winter (no, yes), number of fetuses (1, 2, 3, 4 or more), in vitro fertilization (no, yes), gestational diabetes mellitus (no, yes), overt diabetes mellitus (no, yes), pregnancy-induced hypertension (no, yes), chronic hypertension (no, yes), chorioamnionitis (no, yes), prelabor rupture of membranes (no, yes), prelabor rupture of membranes > 18 h (no, yes), antenatal steroid (no, yes), cesarean section (no, yes), oligohydramnios (no, yes), polysydramnios (no, yes), maternal age (years), primipara (no, yes), maternal education (elementary, junior high, senior high, college or higher), maternal citizenship (Korea, Vietnam, China, Philippines, Japan, Cambodia, United States, Thailand, Mongolia, Other), paternal education (elementary, junior high, senior high, college or higher), paternal citizenship (Korea, Vietnam, China, Philippines, Japan, Cambodia, United States, Thailand, Mongolia, Other), unmarried (no, yes), congenital infection (no, yes), PM10 year (PM10 for each year), PM10 month (PM10 for each birth-month), temperature average (for each year), temperature min (for each year) and temperature max (for each year). PM10 and temperature data came from the Korea Meteorological Administration (PM10 https://data.kma.go.kr/data/climate/SelectDustRltmList.do?pgmNo=68; temperature https://web.kma.go.kr/weather/climate/past_cal.jsp). The definition of each variable is given in Text S1, supplementary text.

Statistical analysis. The artificial neural network, the decision tree, the logistic regression, the Naïve Bayes, the random forest and the support vector machine were used for predicting preterm birth38–43. A decision tree includes three elements, i.e., a test on an independent variable (intermediate note), an outcome of the test
A naïve Bayesian classifier performs classification on the basis of Bayes’ theorem. Here, the theorem states that the probability of the dependent variable given certain values of independent variables can be calculated based on the probabilities of the independent variables given a certain value of the dependent variable. A random forest is a collection of many decision trees, which make majority votes on the dependent variable (“bootstrap aggregation”). Let us take a random forest with 1000 decision trees as an example. Let us assume that original data includes 10,000 participants. Then, the training and test of this random forest takes two steps. Firstly, new data with 10,000 participants is created based on random sampling with replacement, and a decision tree is created based on this new data. Here, some participants in the original data would be excluded from the new data and these leftovers are called out-of-bag data. This process is repeated 1000 times, i.e., 1000 new data are created, 1000 decision trees are created and 1000 out-of-bag data are created. Secondly, the 1000 decision trees make predictions on the dependent variable of every participant in the out-of-bag data, their majority vote is taken as their final prediction on this participant, and the out-of-bag error is calculated as the proportion of wrong votes on all participants in the out-of-bag data38,39.

A support vector machine estimates a group of “support vectors”, that is, a line or space called “hyperplane”. The hyperplane separates data with the greatest gap between various sub-groups. An artificial neural network consists of “neurons”, information units combined through weights. In general, the artificial neural network includes one input layer, one, two or three intermediate layers and one output layer. Neurons in a previous layer link with “weights” in the next layer (Here, these weights denote the strengths of linkages between neurons in a previous layer and their next-layer counterparts). This “feedforward” operation begins from the input layer, runs through intermediate layers and ends in the output layer. Then, this process is followed by learning: These weights are updated according to their contributions for a gap between the actual and predicted final outputs. This “backpropagation” operation begins from the output layer, runs through intermediate layers and ends in the input layer. The two processes are repeated until the performance measure reaches a certain limit38,39. Data on 10,423 observations with full information were divided into training and validation sets with a 70:30 ratio (7296 vs. 3127). Accuracy, a ratio of correct predictions among 3127 observations, was employed as a standard for validating the models. Random forest variable importance, the contribution of a certain variable for the performance (GINI) of the random forest, was used for examining major predictors of adverse birth outcomes in VLBW infants including PM$_{10}$. The random split and analysis were repeated 50 times then its average was taken for external validation44,45. R-Studio 1.3.959 (R-Studio Inc.: Boston, United States) was employed for the analysis during August 1, 2021–September 30, 2021.

---

**Figure 1.** Random forest variable importance plots for adverse birth outcomes: PM$_{10}$ included. PM, particulate matter; PROM, prelabor rupture of membranes.
Table 6. Random forest variable importance rankings for adverse birth outcomes: PM<sub>10</sub> included. The ranking of a top-5 predictor was highlighted with the color of orange and 6–10 predictor was highlighted with the color of mild blue. BW, birth weight (grams); DM, diabetes mellitus; GA, gestational age (weeks); PIH, pregnancy-induced hypertension; PM, particulate matter; PROM, prelabor rupture of membranes; SGA, small-for-gestational age. *P < 0.10 chi-square or T test (Table 2).

| Variable                      | GA<28 | GA<26 | BW<1000 | BW<750 | SGA | Average |
|-------------------------------|-------|-------|---------|--------|-----|---------|
| Sex (Male)                    | 6     | 5     | 5       | 6      | 5   | 4       |
| Birth-Year                    | 16    | 14    | 15      | 14     | 15  | 16      |
| Birth-Month                   | 2     | 2     | 2       | 2      | 2   | 2       |
| Birth-Season: Spring          | 27    | 26    | 27      | 27     | 27  | 27      |
| Birth-Season: Summer          | 23    | 24    | 24      | 22     | 24  | 23      |
| Birth-Season: Autumn          | 24    | 23    | 23      | 23     | 23  | 24      |
| Birth-Season: Winter          | 26    | 25    | 25      | 26     | 25  | 25      |
| Number of fetuses             | 5     | 4     | 7       | 4      | 6   | 5       |
| In vitro fertilization        | 12    | 10    | 9       | 9      | 14  | 12      |
| Gestational DM                | 21    | 22    | 21      | 25     | 21  | 22      |
| Overt DM                      | 33    | 33    | 33      | 33     | 32  | 33      |
| PIH                           | 8     | 8     | 11      | 10     | 4   | 8       |
| Chronic hypertension          | 29    | 29    | 28      | 28     | 28  | 28      |
| Chorioamnionitis              | 4     | 7     | 12      | 15     | 12  | 9       |
| PROM                          | 15    | 15    | 10      | 11     | 7   | 13      |
| PROM > 18 Hour                | 20    | 19    | 20      | 19     | 16  | 20      |
| Antenatal steroid             | 11    | 9     | 13      | 8      | 9   | 10      |
| Cesarean section              | 10    | 11    | 8       | 12     | 13  | 11      |
| Oligohydramnios               | 13    | 13    | 19      | 20     | 8   | 14      |
| Polyhydramnios                | 31    | 31    | 31      | 31     | 30  | 31      |
| Maternal age                  | 1     | 1     | 1       | 1      | 1   | 1       |
| Primipara                     | 7     | 6     | 4       | 5      | 11  | 6       |
| Maternal education            | 9     | 8     | 6       | 7      | 10  | 7       |
| Maternal citizenship          | 25    | 27    | 26      | 24     | 26  | 26      |
| Paternal education            | 14    | 12    | 14      | 13     | 17  | 15      |
| Paternal citizenship          | 28    | 30    | 29      | 29     | 29  | 29      |
| Unmarried                     | 30    | 28    | 30      | 30     | 31  | 30      |
| Congenital infection          | 32    | 32    | 32      | 32     | 33  | 32      |
| PM<sub>10</sub> Year          | 22    | 21    | 22      | 21     | 22  | 21      |
| PM<sub>10</sub> Month         | 3     | 3     | 3       | 3      | 3   | 3       |
| Temperature Year Average      | 18    | 17    | 18      | 18     | 20  | 19      |
| Temperature Year Min          | 19    | 18    | 17      | 17     | 19  | 18      |
| Temperature Year Max          | 17    | 16    | 16      | 16     | 18  | 17      |

Ethic statement. The KNN registry was approved by the institutional review board (IRB) at each participating hospital (IRB No. of Korea University Anam Hospital: 2013AN0115). Informed consent was obtained from the parent(s) of each infant registered in the KNN. All methods were carried out in accordance with the IRB-approved protocol and in compliance with relevant guidelines and regulations. The names of the institutional review board of the KNN participating hospitals were as follows: The institutional review board of Gachon University Gil Medical Center, The Catholic University of Korea Bucheon ST. Mary’s Hospital, The Catholic University of Korea Seoul ST. Mary’s Hospital, The Catholic University of Korea ST. Vincent’s Hospital, The Catholic University of Korea Yeouido ST. Mary’s Hospital, The Catholic University of Korea Uijeongbu ST. Mary’s Hospital, Gangnam Severance Hospital, Kyung Hee University Hospital at Gangdong, GangNeung Asan Hospital, Kangbuk Samsung Hospital, Kangwon National University Hospital, Konkuk University Medical Center, Konyang University Hospital, Kyungpook National University Hospital, Gyeongsang National University Hospital, Kyung Hee University Medical center, Keimyung University Dongsan Medical Center, Korea University Ulsan Cancer Center, Korea University Ansan Hospital, Korea University Anam Hospital, Kosin University Gospel Hospital, National Health Insurance Service Ilsan Hospital, Daegu Catholic University Medical Center, Dongguk University Ilsan Hospital, Dong-A University Hospital, Seoul Metropolitan Government-Seoul National University Boramae Medical Center, Pusan National University Hospital, Busan ST. Mary’s Hospital, Seoul National University Bundang Hospital, Samsung Medical Center, Samsung Changwon Medical Center, Seoul National University Hospital, Asan Medical Center, Sungae Hospital, Severance Hospital, Soonchunhyang University Hospital Bucheon, Soonchunhyang University Hospital Seoul, Soonchunhyang University Hospital Cheonan, Ajou University Hospital, Pusan National University Children’s Hospital, Yeungnam University Hospital, Ulsan University Hospital, Wonkwang University School of Medicine & Hospital, Wondum University Hospital, Eulji University Hospital, Eulji General Hospital, Ewha Womans University Medical.
Center, Inje University Busan Paik Hospital, Inje University Sanggye Paik Hospital, Inje University Ilsan Paik Hospital, Inje University Haedundae Paik Hospital, Inha University Hospital, Chonnam National University Hospital, Chonbuk National University Hospital, Cheil General Hospital & Women's Healthcare Center, Jeju National University Hospital, Chosun University Hospital, Chung-Ang University Hospital, CHA Gangnam Medical Center, CHA University, CHA Bundang Medical Center, CHA University, Chungnam National University Hospital, Chonbuk National University, Kyungpook National University Chilgok Hospital, Kangnam Sacred Heart Hospital, Kangdong Sacred Heart Hospital, Hanyang University Guri Hospital, and Hanyang University Medical Center.

Code and data availability
The code and data used for this study are available from the corresponding author upon reasonable request and under the permission of the Korean Neonatal Network and the Korea Centers for Disease Control and Prevention.

Received: 13 December 2021; Accepted: 6 July 2022
Published online: 01 October 2022

References
1. Liu, L. et al. Global, regional, and national causes of under-5 mortality in 2000–15: An updated systematic analysis with implications for the sustainable development goals. Lancet 388, 3027–3035 (2016).
2. World Health Organization. News: Preterm Birth. http://www.who.int/news-room/factsheets/detail/preterm-birth (2018).
3. Harrison, M. S. & Goldenberg, R. L. Global burden of prematurity. Semin. Fetal Neonatal Med. 21, 74–79 (2016).
4. Bell, E. F. Mortality, in-hospital morbidity, care practices, and 2-year outcomes for extremely preterm infants in the US, 2013–2018. JAMA 327, 248–263 (2022).
5. Johnson, S. & Marlow, N. Early and long-term outcome of infants born extremely preterm. Arch. Dis. Child. 102, 97–102 (2017).
6. Glass, H. C. et al. Outcomes for extremely premature infants. Arch. Dis. Child. 90, 1337–1351 (2015).
7. Holsti, A., Adamsson, M., Hägglöf, B., Farooqi, A. & Serenius, F. Chronic conditions and health care needs of adolescents born at 23 to 25 weeks' gestation. Pediatrics 139, e20162215 (2017).
8. Jensen, E. A. Adverse effects of small for gestational age differ by gestational week among very preterm infants. Arch. Dis. Child. Fetal Neonatal Ed. 104, F192–F198 (2019).
9. Boghosian, N. S., Geraci, M., Edwards, E. M. & Horbar, J. D. Morbidity and mortality in small for gestational age infants at 22 to 29 weeks' gestation. Pediatrics 141, e20172533 (2018).
10. World Health Organization. 2016. Ambient (outdoor) air quality and health. Fact sheet. Retrieved from http://www.who.int/mediacentre/factsheets/fs313/en/ (2016).
11. Lim, S. S. et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: A systematic analysis for the Global Burden of Disease Study 2010. Lancet 380, 2224–2260 (2012).
12. Padula, A. M. et al. Traffic-related air pollution and risk of preterm birth in the San Joaquin Valley of California. Ann. Epidemiol. 24, 888–895 (2014).
13. DeFranco, E. J. et al. Exposure to airborne particulate matter during pregnancy is associated with preterm birth: A population-based cohort study. Environ. Health 15, 6 (2016).
14. Mandola, P. et al. Air pollution and preterm birth: Do air pollution changes over time influence risk in consecutive pregnancies among low-risk women? Int. J. Environ. Res. Public Health 16, 3365 (2019).
15. Lavigne, E. et al. Ambient air pollution and adverse birth outcomes: Differences by maternal comorbidities. Environ. Res. 148, 457–466 (2016).
16. Qian, Z. et al. Ambient air pollution and preterm birth: A prospective birth cohort study in Wuhan, China. Int. J. Hyg. Environ. Health 219, 195–203 (2016).
17. Kim, Y. J. et al. Maternal exposure to particulate matter during pregnancy and adverse birth outcomes in the Republic of Korea. Int. J. Environ. Res. Public Health 16, 633 (2019).
18. Bell, M. L. et al. Prenatal exposure to fine particulate matter and birth weight: Variations by particulate constituents and sources. Epidemiology 21, 884–891 (2010).
19. Fleischer, N. L. et al. Outdoor air pollution, preterm air pollution, and birth weight: Analysis of the world health organization global survey on maternal and perinatal health. Environ. Health Perspect. 122, 425–430 (2014).
20. Lu, C. et al. Combined effects of ambient air pollution and home environmental factors on low birth weight. Chemosphere 240, 124836 (2020).
21. Lu, L. et al. Maternal air pollution exposure increases the risk of preterm birth: Evidence from the meta-analysis of cohort studies. Environ. Res. 202, 111654 (2021).
22. Ye, L. et al. Associations between maternal exposure to air pollution and air pollution and birth outcomes: A retrospective cohort study in Taizhou, China. Environ. Sci. Pollut. Res. Int. 25, 21927–21936 (2018).
23. Liu, W. Y. et al. Association between ambient air pollutants and preterm birth in Ningbo, China: A time-series study. BMC Pediatr. 18, 305 (2018).
24. Trasande, L., Wong, K., Roy, A., Savitz, D. A. & Thurston, G. Exploring prenatal outdoor air pollution, birth outcomes and neonatal health care utilization in a nationally representative sample. J. Expo. Sci. Environ. Epidemiol. 23, 315–321 (2013).
25. Glinianaia, S. V., Rankin, J., Bell, R., Pless-Mulloli, T. & Howel, D. Particulate air pollution and fetal health: A systematic review of the epidemiologic evidence. Epidemiology 15, 36–45 (2004).
26. Backes, C. H., Nelin, T., Gorr, M. W. & Wold, L. E. Early life exposure to air pollution: How bad is it? Toxicol. Lett. 216, 47–53 (2013).
27. Challis, J., Newham, J., Petraglia, F., Yeganeh, M. & Bocking, A. Fetal sex and preterm birth. Placenta 34, 95–99 (2013).
28. Dodd, J. M., Grivell, R. M., O'Brien, C. M., Dowswell, T. & Deussen, A. R. Prenatal administration of prostaglandins for preventing spontaneous preterm birth in women with a multiple pregnancy. Cochrane Database Syst. Rev. 10, CD012024 (2017).
29. Heo, J. S., Lee, H. J., Lee, M. H. & Choi, C. W. Comparison of neonatal outcomes of very low birth weight infants by mode of delivery: In vitro fertilization versus natural pregnancy. Fertil. Steril. 111, 962–970 (2019).
30. Goldenberg, R. L., Cullhane, J. F., Iams, J. D. & Romero, R. Epidemiology and causes of preterm birth. Lancet 371, 75–84 (2008).
31. DiGiulio, D. B. et al. Microbial prevalence, diversity and abundance in amniotic fluid during preterm labor: A molecular and culture-based investigation. PLoS ONE 3, e3056 (2008).
32. Jain, V. G., Willis, K. A., Jobe, A. & Ambalavanan, N. Chorioamnionitis and neonatal outcomes. Pediatr. Res. 91, 289–296 (2022).
33. Sugimoto, H. et al. Neutralization of circulating vascular endothelial growth factor (VEGF) by anti-VEGF antibodies and soluble VEGF receptor 1 (sFlt-1) induces proteinuria. J. Biol. Chem. 278, 12605–12608 (2003).
34. Eremina, V. et al. Glomerular-specific alterations of VEGF-A expression lead to distinct congenital and acquired renal diseases. J. Clin. Invest. 111, 707–716 (2003).
35. Eremina, V. et al. VEGF inhibition and renal thrombotic microangiopathy. N. Engl. J. Med. 358, 1129–1136 (2008).
36. Lu, C. et al. Effect of outdoor air pollution and indoor environmental factors on small for gestational age. Build. Environ. 206, 108399 (2021).
37. Chang, Y. S., Ahn, S. Y. & Park, W. S. The establishment of the Korean Neonatal Network (KNN). Neonatal Med. 20, 169–178 (2013).
38. Lee, K. S. & Ahn, K. H. Artificial neural network analysis of spontaneous preterm labor and birth and its major determinants. J. Korean Med. Sci. 34, e128 (2019).
39. Lee, K. S. & Ahn, K. H. Application of artificial intelligence in early diagnosis of spontaneous preterm labor and birth. Diagnostics (Basel) 10, 733 (2020).
40. Park, E. K. et al. Machine learning approaches to radiogenomics of breast cancer using low-dose perfusion computed tomography: Predicting prognostic biomarkers and molecular subtypes. Sci. Rep. 9, 17847 (2019).
41. Lee, J. Y. et al. Radiomic machine learning for predicting prognostic biomarkers and molecular subtypes of breast cancer using tumor heterogeneity and angiogenesis properties on MRI. Eur. Radiol. 32, 650–660 (2021).
42. Lee, K. S., Song, I. S., Kim, E. S. & Ahn, K. H. Determinants of spontaneous preterm labor and birth including gastroesophageal reflux disease and periodontitis. J. Korean Med. Sci. 35, e105 (2020).
43. Lee, K. S. et al. Association of preterm birth with depression and particulate matter: Machine learning analysis using national health insurance data. Diagnostics (Basel) 11, 555 (2021).
44. Ahn, K. H. et al. Predictors of newborn's weight for height: A machine learning study using nationwide multicenter ultrasound data. Diagnostics (Basel) 11, 1280 (2021).
45. Lee, K. S., Kim, E. S., Kim, D. Y., Song, I. S. & Ahn, K. H. Association of gastroesophageal reflux disease with preterm birth: Machine learning analysis. J. Korean Med. Sci. 36, e282 (2021).

Author contributions
H.C., K.-S.L., E.-H.L. and J.S.H. contributed to the design of the study, the collection, analysis and interpretation of the data, and the writing, editing and approval of the manuscript.

Funding
This work was supported by the Research Program funded by the Korea National Institute of Health (2019-ER7103-02#), the Korea Ministry of Education under the Basic Science Research Program supervised by the National Research Foundation (2020R111A1A01073697), and the Korea Ministry of Science and ICT under the Information Technology Research Center support program supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation) (No. IITP-2018-0-01405). The funder had no role in the design of the study, the collection, analysis and interpretation of the data and the writing of the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-16234-y.

Correspondence and requests for materials should be addressed to K.-S.L. or J.S.H.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2022