

Original Paper

Predicting Fetal Alcohol Spectrum Disorders Using Machine Learning Techniques: Multisite Retrospective Cohort Study

Sarah Soyeon Oh^{1,2}, PhD; Irene Kuang³, PhD; Hyewon Jeong³, MD; Jin-Yeop Song⁴, BSc; Boyu Ren⁵, PhD; Jong Youn Moon⁶, MD, PhD; Eun-Cheol Park², MD, PhD; Ichiro Kawachi¹, MChB, PhD

¹Department of Social and Behavioral Sciences, Harvard TH Chan School of Public Health, Boston, MA, United States

²Institute of Health Services Research, Yonsei University College of Medicine, Seoul, Republic of Korea

³Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, United States

⁴Department of Physics, Massachusetts Institute of Technology, Cambridge, MA, United States

⁵Department of Psychiatry, Harvard Medical School, Boston, MA, United States

⁶Artificial Intelligence and Big-Data Convergence Center, Gil Medical Center, Gachon University College of Medicine, Incheon, Republic of Korea

Corresponding Author:

Jong Youn Moon, MD, PhD

Artificial Intelligence and Big-Data Convergence Center

Gil Medical Center

Gachon University College of Medicine

191 Hambangmoe-ro, Yeonsu-gu

Incheon, 21936

Republic of Korea

Phone: 82 1021245754

Email: moonjy@gachon.ac.kr

Abstract

Background: Fetal alcohol syndrome (FAS) is a lifelong developmental disability that occurs among individuals with prenatal alcohol exposure (PAE). With improved prediction models, FAS can be diagnosed or treated early, if not completely prevented.

Objective: In this study, we sought to compare different machine learning algorithms and their FAS predictive performance among women who consumed alcohol during pregnancy. We also aimed to identify which variables (eg, timing of exposure to alcohol during pregnancy and type of alcohol consumed) were most influential in generating an accurate model.

Methods: Data from the collaborative initiative on fetal alcohol spectrum disorders from 2007 to 2017 were used to gather information about 595 women who consumed alcohol during pregnancy at 5 hospital sites around the United States. To obtain information about PAE, questionnaires or in-person interviews, as well as reviews of medical, legal, or social service records were used to gather information about alcohol consumption. Four different machine learning algorithms (logistic regression, XGBoost, light gradient-boosting machine, and CatBoost) were trained to predict the prevalence of FAS at birth, and model performance was measured by analyzing the area under the receiver operating characteristics curve (AUROC). Of the total cases, 80% were randomly selected for training, while 20% remained as test data sets for predicting FAS. Feature importance was also analyzed using Shapley values for the best-performing algorithm.

Results: Overall, there were 20 cases of FAS within a total population of 595 individuals with PAE. Most of the drinking occurred in the first trimester only (n=491) or throughout all 3 trimesters (n=95); however, there were also reports of drinking in the first and second trimesters only (n=8), and 1 case of drinking in the third trimester only (n=1). The CatBoost method delivered the best performance in terms of AUROC (0.92) and area under the precision-recall curve (AUPRC 0.51), followed by the logistic regression method (AUROC 0.90; AUPRC 0.59), the light gradient-boosting machine (AUROC 0.89; AUPRC 0.52), and XGBoost (AUROC 0.86; AUPRC 0.45). Shapley values in the CatBoost model revealed that 12 variables were considered important in FAS prediction, with drinking throughout all 3 trimesters of pregnancy, maternal age, race, and type of alcoholic beverage consumed (eg, beer, wine, or liquor) scoring highly in overall feature importance. For most predictive measures, the best performance was obtained by the CatBoost algorithm, with an AUROC of 0.92, precision of 0.50, specificity of 0.29, F1 score of 0.29, and accuracy of 0.96.

Conclusions: Machine learning algorithms were able to identify FAS risk with a prediction performance higher than that of previous models among pregnant drinkers. For small training sets, which are common with FAS, boosting mechanisms like CatBoost may help alleviate certain problems associated with data imbalances and difficulties in optimization or generalization.

(*J Med Internet Res* 2023;25:e45041) doi: [10.2196/45041](https://doi.org/10.2196/45041)

KEYWORDS

fetal alcohol syndrome; machine learning; algorithm; development; fetal; fetus; maternal; obstetric; gynecology; pregnant; prenatal; antenatal; postnatal; predict; developmental disability; prenatal alcohol exposure; alcohol; alcohol exposure; developmental; disability; pregnancy; age; race; diagnosis; diagnostic; treatment

Introduction

Fetal alcohol spectrum disorders (FASDs) comprise a range of neuropsychological and behavioral deficits that emerge from prenatal alcohol exposure (PAE) [1]. The most severe form of FASD, otherwise known as fetal alcohol syndrome (FAS), is characterized by distinct facial malformations, prenatal or postnatal growth retardation, and central nervous system abnormalities [2]. In the United States, it is predicted that around 1%-5% of school-aged children have FASDs and 0.6%-0.9% have FAS [3]. While FASDs are 100% preventable if a pregnant woman abstains from consuming alcohol [4], more than 10% of women drink during pregnancy [5]. For certain populations, such as women with alcohol use disorders (AUDs) and women with unintended pregnancies, this rate is higher [6].

Machine learning (ML) proposes an interesting solution to predicting FASDs, as only 1 in every 13 pregnant women who consume alcohol during pregnancy delivers a child with FASDs [7]. However, studies using ML algorithms in FAS detection have been limited to small samples [8] or nonhuman studies [9]. For example, a rodent study of FAS found that certain ML models, including support vector machine-based algorithms of the brain's functional connectivity, have successfully predicted PAE among rodents with an accuracy of up to 62.5%, highlighting the potential for ML-based human subject research [9]. Genome-wide DNA methylation data in small human cohorts (n=48) have achieved moderately accurate predictions of FASD status by using gradient boosting models to distinguish FASD cases and controls [8]. The "K Nearest Neighbor" algorithm for imputing missing prenatal alcohol data has predicted pregnant drinkers with an accuracy of up to 76% and shown potential in imputing missing data for longitudinal studies where data missingness leads to bias [10].

As of now, few studies have attempted to use ML strategies to detect prenatal exposure to alcohol, despite its increased use in retrospective studies of other teratogens like tobacco [11,12], environmental contaminants [13], and certain medications [14]. For example, in an American cohort study of 531 children between 3 and 5 years old, ML algorithms achieved an accuracy of 81% in detecting prenatal exposure to smoking, by incorporating DNA methylation data and maternal self-reports [12]. In another study of longitudinal birth cohorts in New York City, cord blood DNA methylation samples were found to predict average prenatal exposure to air pollution like NO₂ and PM_{2.5} with an accuracy of up to 60% (95% CI 0.52-0.68) [15].

Furthermore, while ML has been used increasingly in recent years to improve the "diagnosis" of FAS—for example, identifying facial features [16]—fewer studies have focused on "predicting" FAS based upon maternal risk factors such as the timing of alcohol exposure during pregnancy (eg, first vs second or third trimester), as well as the frequency and amount of drinking. One reason is because of the difficulty of collecting detailed information about alcohol drinking during pregnancy [9]. With the creation of the collaborative initiative on fetal alcohol spectrum disorders (CIFASDs) in 2003, a consortium of clinicians and researchers from multiple sites in the United States and Europe have begun to collaboratively gather data on prenatal exposure to drugs and alcohol, including data on alcohol exposure histories from maternal reports and review of medical or legal or social service records [17].

As statistical and ML forecasting methods often vary in predictive performance for neonatal studies (eg, ML methods had higher predictive accuracy than traditional statistical methods in predicting mortality among low birthweight infants) [18], this study aims to predict FASDs based on a number of maternal characteristics, and compare or contrast these factors with risk factors highlighted in the existing body of literature where more traditional, statistical methods were used.

Methods

Overview

Data were collected by CIFASD as part of a longitudinal, multisite research study of pregnant drinkers between 2003 and 2017 (for full methodology, see [17]). For this study, data on dysmorphology (U24AA014815), neurobehavior (U01AA014834), and demographics (U01AA014809) were used to gather information about 595 pregnant drinkers who visited 1 of the following sites within the United States to be interviewed about various questions related to their pregnancy behaviors and birth-related outcomes: (1) Center for Behavioral Teratology, San Diego State University, San Diego, CA; (2) Emory University, Atlanta, GA; (3) 7 Northern Plains communities, including 6 Indian reservations; (4) the University of California, Los Angeles, CA; and (5) the University of Minnesota, Minneapolis, MN [17]. Institutional review boards (IRBs) at all CIFASD sites approved this study, and the Harvard T.H. Chan School of Public Health IRB approved analyses of these secondary data (Protocol #: IRB21-1261).

Study Sites

Center for Behavioral Teratology, San Diego, California

At this site, children suspected of alcohol exposure were referred to the principal investigator and local professionals for participation in this project [17]. Many patients were already studying at this center before the initiation of the CIFASD project, including those referred to the investigative team for meeting the traditional diagnostic criteria for FAS (eg, facial anomalies; growth retardation; and evidence of central nervous system dysfunction such as microcephaly, mental retardation, or attentional deficits) [19]. Alcohol exposure histories were obtained via self-reports or professional reviews of medical, legal, or social service records of the biological mother. Parents or primary caregivers completed questionnaires regarding the child's behavior, while the children were examined for facial features of FAS, that is, 2 of the 3 key facial features (short palpebral fissures: ≤ 10 th centile; thin vermilion border of the upper lip: rank 4 or 5 on a racially normed lip or philtrum guide; smooth philtrum: rank 4 or 5 on a racially normed lip or philtrum guide), as well as signs of prenatal or postnatal growth deficiency (head circumference or height or weight ≤ 10 th percentile) [20].

Emory University, Atlanta, Georgia

At Emory University, the Fetal Alcohol and Drug Exposure Clinic gathered data on a large sample of patients with FASDs while providing clinical services and facial evaluations at the Emory University Marcus Institute [21]. In the absence of direct reports, documentation of alcohol abuse or dependence by the biological mother in the form of medical, social services, or court records was reviewed [17]. Recruitment took place via clinical and community referrals, and parents or primary caregivers completed questionnaires or interviews, while patients with FASD were administered various neuropsychological tests over a 3-hour session [22].

Northern Plains

Seven communities, including 1 urban and 6 reservation sites throughout North Dakota, South Dakota, and Montana, participated in this study [21]. Children with FASDs were recruited via active case ascertainment methods and advertisements in tribal and community health centers [20]. Data on PAE were obtained from in-person interviews with the parent or primary caregiver to obtain exact exposure histories retrospectively and were also confirmed via reviews of medical records, when available [20].

University of Minnesota, Minneapolis, Minnesota

The Department of Psychiatry at the University of Minnesota collected data on PAE histories obtained through several modalities including medical reports, birth records, social service records, and when available, using maternal self-reports [21].

University of California, Los Angeles

Data were collected from children attending the Fetal Alcohol and Related Disorders Clinic at University of California, Los Angeles (UCLA) [23]. Participant recruitment was through local FASD clinic referrals, web-based advertisements, and word of mouth in caregiving communities [23]. All alcohol

exposure histories were confirmed via in-person interviews, maternal reports of prenatal substance exposure, or the review of maternal medical records by a licensed medical doctor [23].

To obtain information about PAE, questionnaires or in-person interviews as well as reviews of medical, legal, or social service records regarding alcohol-related problems or a diagnosis of alcohol abuse were used to gather information about alcohol consumption. At all CIFASD sites, the Institute of Medicine's definition of FASDs was used for diagnosis, for example, (1) evidence of a characteristic pattern of minor facial anomalies including at least 2 or more of the key facial features of FAS (palpebral fissures ≤ 10 th centile, thin vermilion border, or smooth philtrum), (2) evidence of prenatal and postnatal growth retardation (height or weight ≤ 10 th centile), (3) evidence of deficient brain growth (structural brain anomalies or occipitofrontal circumference ≤ 10 th centile), and if possible (4), confirmation of maternal alcohol consumption directly from the mother or a knowledgeable collateral source was used for FAS diagnosis [2] were used by dysmorphologists at each site to diagnose FAS. Among children with confirmed PAE, a diagnosis of FAS was made if 2 of the 3 key facial features of FAS (ie, short palpebral fissure, smooth philtrum, or thin vermilion) was accompanied by either microcephaly, growth retardation, or both [17]. Children were excluded when there were reports of known causes of mental deficiency, such as congenital hypothyroidism, neurofibromatosis, or chromosomal abnormalities.

At 4 of the sites including Emory University, the University of California, University of Minnesota, and San Diego, a dysmorphologist trained to accurately diagnose FAS based on physical features, as defined by the CIFASD Dysmorphology Core, was used to diagnose FAS. Contrastingly, for the Northern Plains site, a team of physicians, teachers, and other representatives were trained to identify children with certain morphological characteristics of FASD and other birth defects, IQ, and neuropsychologic traits; however, subjects could not always be referred to a pediatric dysmorphologist for verification or a complete physical examination or morphology assessment; resulting in difficulties with ethnic variations in morphology, syndromic features of FASDs were sometimes compared to normal controls within the same population, in terms of weight, head circumference, fissure length, and other facial characteristics (eg, ptosis and intercanthal distance) [24].

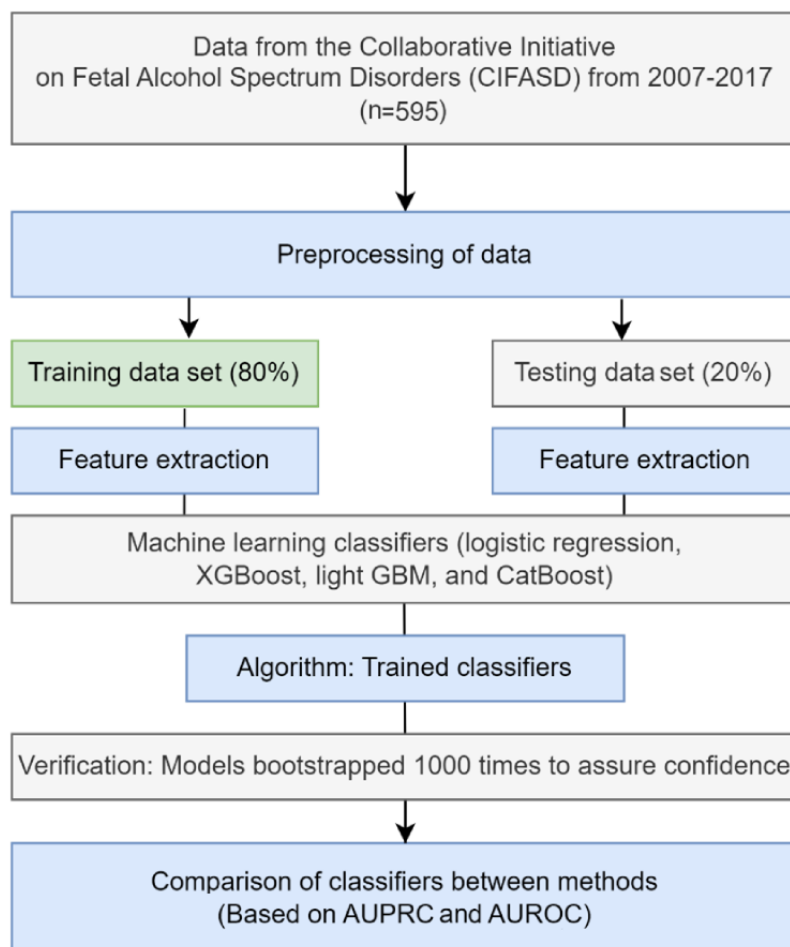
Regarding PAE, all cases of biological mothers with reported alcohol consumption during pregnancy were categorized into 1 of the following mutually exclusive groups: (1) women who consumed alcohol in the first trimester only, (2) women who consumed alcohol in the first and second trimesters only, (3) women who consumed alcohol during all 3 trimesters of pregnancy, and (4) women with "other" drinking patterns (eg, drinking only in the second or third trimester). Preferred alcoholic beverage type (eg, "beer," "wine," and "spirits"), maternal age, maternal race (American Indian or Alaska Native, Asian, Native Hawaiian or other Pacific Islander, Black or African American, White, more than 1 race, other), ethnicity (Hispanic or Latino, or Not Hispanic or Latino), the reception of prenatal care (yes, no), and experience of any

pregnancy-related complications (bleeding, high blood pressure, diabetes) were also included in all models.

For the computational analyses, 80% ($n=595$) of the total cases were randomly selected for training, while 20% remained as test data sets for predicting FAS (Figure 1). The measures for the predictive performance of each algorithm (logistic regression, XGBoost, light gradient-boosting machine [GBM], and CatBoost) included area under the precision-recall curve (AUPRC) and area under the receiver operating characteristic

curve (AUROC). Besides logistic regression, which has been the traditional approach for making associations between pregnant drinking patterns and FASDs in the existing body of literature [25,26], on boosting (XGBoost [27], light GBM [28], CatBoost [29]), a supervised ML method that consists of aggregating classifiers developed sequentially on the train-test sample, to learn from classifiers, correct errors, and obtain more accurate classifiers by training a sequence of weaker models [30].

Figure 1. Flow diagram of investigation. AUPRC: area under the precision-recall curve; AUROC: receiver operating characteristics curve; GBM: gradient-boosting machine.



In order to verify the prediction performance of our 4 methods, the performance of each of the 4 models was bootstrapped 1000 times for our entire sample ($n=595$) and validated. A box plot of the mean and pooled standard deviations of each model over 1000 bootstrapped samples can be found in [Multimedia Appendix 1](#). All models had a similar range in variance with pooled standard deviations ranging from 0.07 (light GBM) to 0.09 (CatBoost, XG Boost, logistic regression). As CatBoost continued to show the highest prediction performance out of all 4 models following verification (mean 0.94, SD 0.09), the importance of features based on Shapley values was verified based on the CatBoost model, and Shapley additive explanation values [31] were used to interpret how each feature contributed to the prediction of FAS risk based on the CatBoost model. All statistical analyses were performed using Python (version 3.7.2; Python Software Foundation) and Scikit-learn library (version 0.20.2; David Cournapeau and Matthieu Brucher) [32].

Ethics Approval

This study was approved by the IRB at the Harvard TH Chan School of Public Health (IRB approval number: IRB21-1261), and the procedures were conducted in accordance with the Helsinki Declaration of 1975, as revised in 2000 for human subjects research. During primary data collection at each clinical location of CIFASD, IRB approval and informed consent were obtained from all adult participants or their legal guardians [33]. For secondary analysis of the data, our research team was provided with deidentified and anonymized data upon request and approval from CIFASD's data committee.

Results

Table 1 presents the data characteristics of the study population. Overall, there were 20 cases of FAS within a total population

of 595 individuals with PAE. Most of the drinking occurred in the first trimester only (n=491) or throughout all 3 trimesters (n=95); however, there were also reports of drinking in the first and second trimesters only (n=8), and 1 case of drinking in the third trimester only (n=1).

Figure 2 presents the AUROC and the AUPRC of each ML algorithm. The CatBoost method delivered the best performance in terms of AUROC (0.92) and AUPRC (0.51), followed by the logistic regression method (AUROC 0.90; AUPRC 0.59), light GBM (AUROC 0.89; AUPRC 0.52), and XGBoost (AUROC 0.86; AUPRC 0.45).

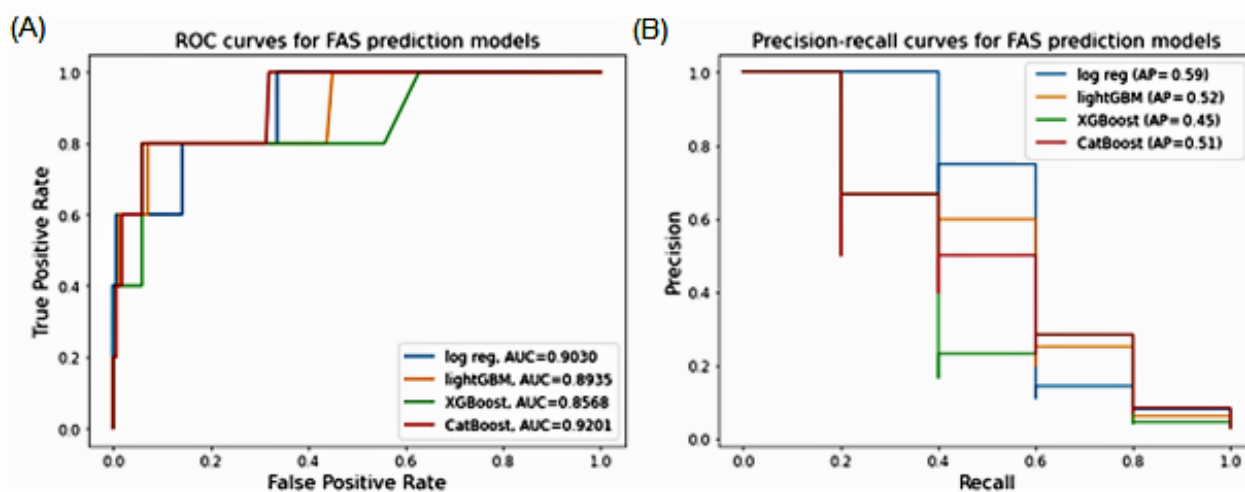
Table 1. Data characteristics of study population.

Characteristics	Total	FAS ^a (n=20)	No FAS (n=575)
Drinking timing, n (%)			
First trimester only	491	3 (0.6)	488 (99.4)
First and second trimesters	8	0 (0.0)	8 (100.0)
All throughout	95	17 (17.9)	78 (82.1)
Other	1	0 (0.0)	1 (100.0)
Race, n (%)			
American Indian or Alaska Native	79	1 (1.3)	78 (98.7)
Black	238	4 (1.7)	234 (98.3)
White	237	10 (4.2)	227 (95.8)
Other ^b	41	5 (12.2)	36 (87.8)
Ethnicity, n (%)			
Hispanic or Latino	92	2 (2.2)	90 (97.8)
Not Hispanic or Latino	503	18 (3.6)	485 (96.4)
Maternal age at childbirth, mean (SD)	25.45 (3.59)	24.30 (2.93)	25.93 (3.61)
Preferred alcoholic beverage, n (%)			
Beer	73	12 (16.4)	61 (83.6)
Wine	496	3 (0.6)	493 (99.4)
Liquor	26	5 (19.2)	21 (80.8)
Prenatal care, n (%)			
No	45	5 (11.1)	40 (88.9)
Yes	550	15 (2.7)	535 (97.3)
Pregnancy complications, n (%)			
No	468	16 (3.4)	452 (96.6)
Yes	127	4 (3.1)	123 (96.9)
Total, n (%)	595	20 (3.4)	575 (96.6)

^aFAS: fetal alcohol syndrome.

^bAsian, Native Hawaiian or other Pacific Islander, or more than 1 race.

Figure 2. Performance evaluation of machine learning (ML) algorithms for FAS prediction in the collaborative initiative on fetal alcohol spectrum disorder (CIFASD) data set. AUC: area under the receiver operating characteristic curve; AP: average precision; FAS: fetal alcohol syndrome; GBM: gradient-boosting machine; ROC: receiver operating characteristic.



Shapley values illustrated that drinking throughout all 3 trimesters of pregnancy, maternal age, race, and type of alcoholic beverage consumed were observed to be the most important features for prediction (mean 1.08), followed by maternal age (mean 0.79), race (mean 0.52), beverage type (mean 0.4), pregnancy complications (mean 0.2), ethnicity (mean 0.07), and prenatal care (mean 0.03). Regarding alcohol consumption amount, drinking throughout all 3 trimesters was associated with FAS risk, however, only drinking during the first trimester, drinking during the first and second trimesters, and other patterns were not associated with increased FAS risk.

Discussion

Principal Findings

FAS is more common among socioeconomically disadvantaged populations but is also more likely to be underdiagnosed due to inadequate resources [34]. The significance of our research from an algorithm development perspective is that our study is the first of its kind to use ML algorithms to predict FASD onset, via the incorporation of variables associated with maternal pregnancy behaviors and the CIFASD data set. Despite FASDs being 100% preventable in nature, the application of automated methods such as ML for the identification of high-risk groups remains rare, relative to other neurodevelopmental disorders.

However, while few FASD studies have incorporated ML for “disease prediction” based on maternal behaviors and sociodemographic characteristics, a growing body of literature has been incorporating ML for “diagnostic purposes” using facial features’ data to distinguish FASD children from non-FASD children [35]. While this was not the objective of our study, it is interesting to note that recent studies amalgamating ML methods such as decision trees, support vector machine, and k-nearest neighbor with 3D-metric facial data for FAS diagnosis have been able to achieve an accuracy rate of up to 89% in clinical settings [35]. Scholars have remarked that advances in FASD diagnosis are often “hindered by a lack of consensus in diagnostic criteria and limited use of objective biomarkers,” highlighting the value of such studies to aid clinical decision-making.

For other intellectual disabilities with larger data sets like autism spectrum disorder (ASD) and epilepsy [36], accuracy rates have ranged from 72.40% [37] to 86.64% [38]. However, in 1 study of children with ASDs, data sets incorporating graph signal processing data were able to reach a diagnostic accuracy of up to 100% in differentiating ASD patients from typically developing children [39]. These studies suggest that for neurodevelopmental illnesses like ASD, artificial intelligence techniques could aid physicians to apply automatic diagnosis and rehabilitation procedures with great accuracy in the future [40].

In 2 of our ML models, the use of only 12 variables centralized around self-reportable measures of alcohol consumption during pregnancy and basic sociodemographic characteristics (age and race or ethnicity) resulted in a predictive accuracy of over 90%. While it may be common knowledge that drinking any amount of alcohol can harm the fetus, it is important to understand that information regarding dose, timing, type, and frequency can improve the prediction of FAS.

Comparison to Prior Work

Regarding ML, scholars have emphasized that there are numerous issues with interpretability and inference, including overgeneralization or overinterpretation of causality [35]. Likewise, because our data set was small and FASD prevalence was low for numerous scenarios, certain confusion matrices had higher numbers for identifying true negatives than true positives, resulting in substantial imprecision in the estimates of “sensitivity” and “precision” [36]. While this may be common among rare outcomes, it highlights the need to gather more data on pregnant drinkers in future studies.

In our study, we also evaluated the AUPRC as a performance metric for FASD. Precision-recall curves are based on precision rather than the false-positive rate and are noted by scholars to be a better assessment of model performance when predicting outcomes that are rare or “unbalanced” due to a small data set [37]. While our curve was beneficial in helping understand the magnitude and uncertainty of each ML algorithm’s performance, in a real-world setting, skewed class distribution will likely be

inevitable [38]. Thus, we recommend that instead of using ML algorithms alone, these models should be used in combination with other, externally validating screening or surveillance strategies to identify high-risk FASD groups, for example, women with a history of alcohol abuse during pregnancy or multiple children with FASDs [39].

Strengths and Limitations

Our model only incorporates a small number of variables that could easily be collected by health systems via self-registrable questionnaires or routine data collection methods from health records [40]. This may be beneficial as some skepticism has been expressed regarding the feasibility of implementing ML models in real-life health care practices [41]. Among the different algorithms tested, the CatBoost algorithm had the highest predictive performance of all algorithms. As noted previously by other researchers, because CatBoost is the newest gradient-boosting decision tree algorithm with better handling of categorical features compared to other algorithms, it will usually outperform other models such as XGBoost and light GBM [42].

This study has several limitations. First, like all traditional studies of FASDs, some of the self-reported data on PAE is likely unreliable and influenced by social desirability or retrospective recall bias [43]. For example, it has been noted that pregnant women often present a more favorable image of themselves when it comes to self-reporting questionnaires about their dietary intake during pregnancy [44]. As a solution, researchers are increasingly exploring the use of biospecimens including meconium, urine, the placenta, neonatal blood,

maternal blood, and fetal tissue (ie, the umbilical cord) to extract biomarkers like fatty acid ethyl esters, ethyl glucuronide, ethyl sulfate, and phosphatidylethanol to detect PAE [45]. Second, certain barriers to data collection were inevitable because of barriers such as a limited window of detectability, difficulties in collection, and high costs of analysis [45]. Such barriers may be overcome in future FASD studies if data on biomarkers are used in combination with maternal self-reporting to improve prediction. Third, besides logistic regression, our study mainly focused on boosting algorithms because they are known to reduce variance and have higher flexibility or interpretability in ML ensembles [46]. However, for training sets that are small like ours, boosting mechanisms may formulate discriminative classifiers where the optimality criterion that the loss function approximates is unclear [47]. Thus, future studies would benefit from incorporating other algorithms that are not boosting-based, such as random forest, for a more well-rounded analysis.

Conclusions

In this study comparing multiple ML algorithms to predict FAS risk among a sample of pregnant drinkers, the CatBoost model outperformed both traditional and other ML models. The variables and methods used in our CatBoost model may serve as an effective, automated method for identifying high-risk groups in clinical predictions of FAS. Future research should evaluate the accuracy of such methods in predicting FAS relative to traditional approaches such as logistic regression analysis, as well as the extent to which certain risk factors may have been missed or overlooked, for overall improved clinical outcomes among FAS patients.

Acknowledgments

This research was supported by a grant of the Korea Health Promotion R&D Project, funded by the Ministry of Health & Welfare, Republic of Korea (grant number HS21C0037). We thank the National Institute on Alcohol Abuse and Alcoholism Collaborative Initiative on Fetal Alcohol Spectrum Disorder for providing us with the dysmorphology (U24AA014815), neurobehavior (U01AA014834), and management (U01AA014809) data sets used in this project. We thank Leah Wetherill for providing us with her insight and guidance. All data were obtained from [48].

Data Availability

Access to the National Institutes of Health's Collaborative Initiative on Fetal Alcohol Spectrum Disorder (CIFASD) data set can be requested from [48]. The CIFASD Data Access Committee will review and make available archived data for all research that aims to improve diagnoses, interventions, and treatment of fetal alcohol syndrome disorders.

Authors' Contributions

SSO analyzed the data and wrote the original draft. I Kuang analyzed the data and performed mock peer review. JYS coordinated data bootstrapping and coding. HJ performed mock peer review and helped write the discussion. BR provided biostatistical consulting. JYM acquired funding, provided mock peer review, and supervised the investigation. ECP provided mock peer review and supervised the investigation. I Kawachi wrote the original draft, provided mock peer review, and supervised the investigation.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Box plot showing area under the receiver operating characteristic curve variation among 4 models. GBM: gradient-boosting machine.

[[PNG File , 19 KB-Multimedia Appendix 1](#)]

References

1. Burns L, Breen C, Bower C, O' Leary C, Elliott EJ. Counting fetal alcohol spectrum disorder in Australia: the evidence and the challenges. *Drug Alcohol Rev* 2013;32(5):461-467 [doi: [10.1111/dar.12047](https://doi.org/10.1111/dar.12047)] [Medline: [23617437](https://pubmed.ncbi.nlm.nih.gov/23617437/)]
2. Hoyme HE, May PA, Kalberg WO, Kodituwakku P, Gossage JP, Trujillo PM, et al. A practical clinical approach to diagnosis of fetal alcohol spectrum disorders: clarification of the 1996 institute of medicine criteria. *Pediatrics* 2005;115(1):39-47 [FREE Full text] [doi: [10.1542/peds.2004-0259](https://doi.org/10.1542/peds.2004-0259)] [Medline: [15629980](https://pubmed.ncbi.nlm.nih.gov/15629980/)]
3. May PA, Baete A, Russo J, Elliott AJ, Blankenship J, Kalberg WO, et al. Prevalence and characteristics of fetal alcohol spectrum disorders. *Pediatrics* 2014;134(5):855-866 [FREE Full text] [doi: [10.1542/peds.2013-3319](https://doi.org/10.1542/peds.2013-3319)] [Medline: [25349310](https://pubmed.ncbi.nlm.nih.gov/25349310/)]
4. Senturias YSN. Fetal alcohol spectrum disorders: an overview for pediatric and adolescent care providers. *Curr Probl Pediatr Adolesc Health Care* 2014;44(4):74-81 [doi: [10.1016/j.cppeds.2013.12.012](https://doi.org/10.1016/j.cppeds.2013.12.012)] [Medline: [24810409](https://pubmed.ncbi.nlm.nih.gov/24810409/)]
5. Burns E, Gray R, Smith LA. Brief screening questionnaires to identify problem drinking during pregnancy: a systematic review. *Addiction* 2010;105(4):601-614 [doi: [10.1111/j.1360-0443.2009.02842.x](https://doi.org/10.1111/j.1360-0443.2009.02842.x)] [Medline: [20403013](https://pubmed.ncbi.nlm.nih.gov/20403013/)]
6. Ethen MK, Ramadhani TA, Scheuerle AE, Canfield MA, Wyszynski DF, Druschel CM, National Birth Defects Prevention Study. Alcohol consumption by women before and during pregnancy. *Matern Child Health J* 2009;13(2):274-285 [FREE Full text] [doi: [10.1007/s10995-008-0328-2](https://doi.org/10.1007/s10995-008-0328-2)] [Medline: [18317893](https://pubmed.ncbi.nlm.nih.gov/18317893/)]
7. Lange S, Probst C, Gmel G, Rehm J, Burd L, Popova S. Global prevalence of fetal alcohol spectrum disorder among children and youth: a systematic review and meta-analysis. *JAMA Pediatr* 2017 01;171(10):948-956 [FREE Full text] [doi: [10.1001/jamapediatrics.2017.1919](https://doi.org/10.1001/jamapediatrics.2017.1919)] [Medline: [28828483](https://pubmed.ncbi.nlm.nih.gov/28828483/)]
8. Lussier AA, Morin AM, MacIsaac JL, Salmon J, Weinberg J, Reynolds JN, et al. DNA methylation as a predictor of fetal alcohol spectrum disorder. *Clin Epigenetics* 2018;10:5 [FREE Full text] [doi: [10.1186/s13148-018-0439-6](https://doi.org/10.1186/s13148-018-0439-6)] [Medline: [29344313](https://pubmed.ncbi.nlm.nih.gov/29344313/)]
9. Rodriguez CI, Vergara VM, Davies S, Calhoun VD, Savage DD, Hamilton DA. Detection of prenatal alcohol exposure using machine learning classification of resting-state functional network connectivity data. *Alcohol* 2021;93:25-34 [doi: [10.1016/j.alcohol.2021.03.001](https://doi.org/10.1016/j.alcohol.2021.03.001)] [Medline: [33716098](https://pubmed.ncbi.nlm.nih.gov/33716098/)]
10. Sania A, Pini N, Nelson M, Myers MM, Shuffrey LC, Lucchini M, et al. The K nearest neighbor algorithm for imputation of missing longitudinal prenatal alcohol data. SSRN. Preprint posted online on March 24, 2022 [FREE Full text] [doi: [10.21203/rs.3.rs-32456/v2](https://doi.org/10.21203/rs.3.rs-32456/v2)]
11. Joubert BR, Håberg SE, Nilsen RM, Wang X, Vollset SE, Murphy SK, et al. 450K epigenome-wide scan identifies differential DNA methylation in newborns related to maternal smoking during pregnancy. *Environ Health Perspect* 2012;120(10):1425-1431 [FREE Full text] [doi: [10.1289/ehp.1205412](https://doi.org/10.1289/ehp.1205412)] [Medline: [22851337](https://pubmed.ncbi.nlm.nih.gov/22851337/)]
12. Ladd-Acosta C, Shu C, Lee BK, Gidaya N, Singer A, Schieve LA, et al. Presence of an epigenetic signature of prenatal cigarette smoke exposure in childhood. *Environ Res* 2016;144(Pt A):139-148 [FREE Full text] [doi: [10.1016/j.envres.2015.11.014](https://doi.org/10.1016/j.envres.2015.11.014)] [Medline: [26610292](https://pubmed.ncbi.nlm.nih.gov/26610292/)]
13. Mobarak YM. Review of the developmental toxicity and teratogenicity of three environmental contaminants (cadmium, lead and mercury). *Catrina: Int J Environ Sci* 2008;3(1):31-43 [FREE Full text]
14. Kang L, Duan Y, Chen C, Li S, Li M, Chen L, et al. Structure-activity relationship (SAR) model for predicting teratogenic risk of antiepileptic medications in pregnancy by using support vector machine. *Front Pharmacol* 2022;13:747935 [FREE Full text] [doi: [10.3389/fphar.2022.747935](https://doi.org/10.3389/fphar.2022.747935)] [Medline: [35281912](https://pubmed.ncbi.nlm.nih.gov/35281912/)]
15. Wang Y, Perera F, Guo J, Riley KW, Durham T, Ross Z, et al. A methodological pipeline to generate an epigenetic marker of prenatal exposure to air pollution indicators. *Epigenetics* 2022;17(1):32-40 [FREE Full text] [doi: [10.1080/15592294.2021.1872926](https://doi.org/10.1080/15592294.2021.1872926)] [Medline: [33465004](https://pubmed.ncbi.nlm.nih.gov/33465004/)]
16. Zhang C, Itti L, Tseng PH, Paolozza A, Reynolds JN, Munoz DP. Machine learning-based screening for fetal alcohol spectrum disorder. In: Proceedings of the AI for Social Good NeurIPS Workshop. 2018 Presented at: AI for Social Good NeurIPS Workshop; December 2018; Montreal, QC p. 3-8
17. Mattson SN, Foroud T, Sowell ER, Jones KL, Coles CD, Fagerlund A, CIFASD. Collaborative initiative on fetal alcohol spectrum disorders: methodology of clinical projects. *Alcohol* 2010;44(7-8):635-641 [FREE Full text] [doi: [10.1016/j.alcohol.2009.08.005](https://doi.org/10.1016/j.alcohol.2009.08.005)] [Medline: [20036488](https://pubmed.ncbi.nlm.nih.gov/20036488/)]
18. Do HJ, Moon KM, Jin HS. Machine learning models for predicting mortality in 7472 very low birth weight infants using data from a nationwide neonatal network. *Diagnostics (Basel)* 2022;12(3):625 [FREE Full text] [doi: [10.3390/diagnostics12030625](https://doi.org/10.3390/diagnostics12030625)] [Medline: [35328178](https://pubmed.ncbi.nlm.nih.gov/35328178/)]
19. Doyle LR, Coles CD, Kable JA, May PA, Sowell ER, Jones KL, CIFASD. Relation between adaptive function and IQ among youth with histories of heavy prenatal alcohol exposure. *Birth Defects Res* 2019;111(12):812-821 [FREE Full text] [doi: [10.1002/bdr2.1463](https://doi.org/10.1002/bdr2.1463)] [Medline: [30719847](https://pubmed.ncbi.nlm.nih.gov/30719847/)]
20. Bernes GA, Courchesne-Krak NS, Hyland MT, Villodas MT, Coles CD, Kable JA, CIFASD. Development and validation of a postnatal risk score that identifies children with prenatal alcohol exposure. *Alcohol Clin Exp Res* 2022;46(1):52-65 [FREE Full text] [doi: [10.1111/acer.14749](https://doi.org/10.1111/acer.14749)] [Medline: [34806190](https://pubmed.ncbi.nlm.nih.gov/34806190/)]

21. Suttie M, Wetherill L, Jacobson SW, Jacobson JL, Hoyme HE, Sowell ER, CIFASD. Facial curvature detects and explicates ethnic differences in effects of prenatal alcohol exposure. *Alcohol Clin Exp Res* 2017;41(8):1471-1483 [FREE Full text] [doi: [10.1111/acer.13429](https://doi.org/10.1111/acer.13429)] [Medline: [28608920](https://pubmed.ncbi.nlm.nih.gov/28608920/)]
22. Doyle LR, Glass L, Wozniak JR, Kable JA, Riley EP, Coles CD, CIFASD. Relation between oppositional/conduct behaviors and executive function among youth with histories of heavy prenatal alcohol exposure. *Alcohol Clin Exp Res* 2019;43(6):1135-1144 [doi: [10.1111/acer.14036](https://doi.org/10.1111/acer.14036)] [Medline: [30908651](https://pubmed.ncbi.nlm.nih.gov/30908651/)]
23. Mattson SN, Roesch SC, Fagerlund A, Autti-Rämö I, Jones KL, May PA, Collaborative Initiative on Fetal Alcohol Spectrum Disorders (CIFASD). Toward a neurobehavioral profile of fetal alcohol spectrum disorders. *Alcohol Clin Exp Res* 2010;34(9):1640-1650 [FREE Full text] [doi: [10.1111/j.1530-0277.2010.01250.x](https://doi.org/10.1111/j.1530-0277.2010.01250.x)] [Medline: [20569243](https://pubmed.ncbi.nlm.nih.gov/20569243/)]
24. May PA, Gossage JP, Smith M, Tabachnick BG, Robinson LK, Manning M, et al. Population differences in dysmorphic features among children with fetal alcohol spectrum disorders. *J Dev Behav Pediatr* 2010;31(4):304-316 [FREE Full text] [doi: [10.1097/DBP.0b013e3181dae243](https://doi.org/10.1097/DBP.0b013e3181dae243)] [Medline: [20431397](https://pubmed.ncbi.nlm.nih.gov/20431397/)]
25. May PA, Blankenship J, Marais AS, Gossage JP, Kalberg WO, Joubert B, et al. Maternal alcohol consumption producing fetal alcohol spectrum disorders (FASD): quantity, frequency, and timing of drinking. *Drug Alcohol Depend* 2013;133(2):502-512 [FREE Full text] [doi: [10.1016/j.drugalcdep.2013.07.013](https://doi.org/10.1016/j.drugalcdep.2013.07.013)] [Medline: [23932841](https://pubmed.ncbi.nlm.nih.gov/23932841/)]
26. O'Leary CM, Nassar N, Kurinczuk JJ, de Klerk N, Geelhoed E, Elliott EJ, et al. Prenatal alcohol exposure and risk of birth defects. *Pediatrics* 2010;126(4):e843-e850 [doi: [10.1542/peds.2010-0256](https://doi.org/10.1542/peds.2010-0256)] [Medline: [20876169](https://pubmed.ncbi.nlm.nih.gov/20876169/)]
27. Chen T, Guestrin C. Xgboost: a scalable tree boosting system. Xgboost: A scalable tree boosting system. Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining; 2016 Presented at: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 13-17, 2016; San Francisco, CA, USA p. 785-794 [doi: [10.1145/2939672.2939785](https://doi.org/10.1145/2939672.2939785)]
28. 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;30:1-9 [FREE Full text]
29. Dorogush AV, Ershov V, Gulin A. CatBoost: gradient boosting with categorical features support. ArXiv. Preprint posted online on October 24, 2018 [FREE Full text]
30. Hamim T, Benabbou F, Sael N. Student profile modeling using boosting algorithms. *Int J Web-Based Learn Teach Technol* 2022;17(5):1-13 [doi: [10.4018/ijwltt.20220901.0a4](https://doi.org/10.4018/ijwltt.20220901.0a4)]
31. Lundberg S, Lee SI. A unified approach to interpreting model predictions. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017 Presented at: NIPS'17; December 4-9, 2017; Long Beach, CA p. 4768-4777
32. Varoquaux G, Buitinck L, Louppe G, Grisel O, Pedregosa F, Mueller A. Scikit-learn. *GetMobile Mobile Comp Comm* 2015 Jun;19(1):29-33 [doi: [10.1145/2786984.2786995](https://doi.org/10.1145/2786984.2786995)]
33. Kable JA, Mehta PK, Rashid F, Coles CD. Path analysis of the impact of prenatal alcohol on adult vascular function. *Alcohol (Hanover)* 2023;47(1):116-126 [doi: [10.1111/acer.14970](https://doi.org/10.1111/acer.14970)] [Medline: [36330744](https://pubmed.ncbi.nlm.nih.gov/36330744/)]
34. Wozniak JR, Riley EP, Charness ME. Clinical presentation, diagnosis, and management of fetal alcohol spectrum disorder. *Lancet Neurol* 2019;18(8):760-770 [FREE Full text] [doi: [10.1016/S1474-4422\(19\)30150-4](https://doi.org/10.1016/S1474-4422(19)30150-4)] [Medline: [31160204](https://pubmed.ncbi.nlm.nih.gov/31160204/)]
35. Lau N, Fridman L, Borghetti BJ, Lee JD. Machine learning and human factors: status, applications, and future directions. *Proc Hum Factors Ergon Soc* 2018;62(1):135-138 [doi: [10.1177/1541931218621031](https://doi.org/10.1177/1541931218621031)]
36. Gaynes BN, Gavin N, Meltzer-Brody S, Lohr KN, Swinson T, Gartlehner G, et al. Perinatal depression: prevalence, screening accuracy, and screening outcomes. *Evid Rep Technol Assess (Summ)* 2005(119):1-8 [doi: [10.1037/e439372005-001](https://doi.org/10.1037/e439372005-001)] [Medline: [15760246](https://pubmed.ncbi.nlm.nih.gov/15760246/)]
37. Sofaer HR, Hoeting JA, Jarnevich CS. The area under the precision - recall curve as a performance metric for rare binary events. *Methods Ecol Evol* 2019;10(4):565-577 [FREE Full text] [doi: [10.1111/2041-210x.13140](https://doi.org/10.1111/2041-210x.13140)]
38. Miao J, Zhu W. Precision-recall curve (PRC) classification trees. *Evol Intel* 2022;15(3):1545-1569 [doi: [10.1007/s12065-021-00565-2](https://doi.org/10.1007/s12065-021-00565-2)]
39. May PA, Gossage JP. Maternal risk factors for fetal alcohol spectrum disorders: not as simple as it might seem. *Alcohol Res Health* 2011;34(1):15-26 [FREE Full text] [Medline: [23580036](https://pubmed.ncbi.nlm.nih.gov/23580036/)]
40. 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(1):322 [FREE Full text] [doi: [10.1186/s12887-021-02788-9](https://doi.org/10.1186/s12887-021-02788-9)] [Medline: [34289819](https://pubmed.ncbi.nlm.nih.gov/34289819/)]
41. Asan O, Choudhury A. Research trends in artificial intelligence applications in human factors health care: mapping review. *JMIR Hum Factors* 2021 Jun 18;8(2):e28236 [FREE Full text] [doi: [10.2196/28236](https://doi.org/10.2196/28236)] [Medline: [34142968](https://pubmed.ncbi.nlm.nih.gov/34142968/)]
42. Hancock JT, Khoshgoftaar TM. CatBoost for big data: an interdisciplinary review. *J Big Data* 2020;7(1):94 [FREE Full text] [doi: [10.1186/s40537-020-00369-8](https://doi.org/10.1186/s40537-020-00369-8)] [Medline: [33169094](https://pubmed.ncbi.nlm.nih.gov/33169094/)]
43. Eichler A, Grunitz J, Grimm J, Walz L, Raabe E, Goecke TW, et al. Did you drink alcohol during pregnancy? Inaccuracy and discontinuity of women's self-reports: on the way to establish meconium ethyl glucuronide (EtG) as a biomarker for alcohol consumption during pregnancy. *Alcohol* 2016;54:39-44 [doi: [10.1016/j.alcohol.2016.07.002](https://doi.org/10.1016/j.alcohol.2016.07.002)] [Medline: [27565755](https://pubmed.ncbi.nlm.nih.gov/27565755/)]
44. van de Mortel TF. Faking it: social desirability response bias in self-report research. *Aust J Adv Nurs* 2008;25(4):40-48 [FREE Full text]

45. Bager H, Christensen LP, Husby S, Bjerregaard L. Biomarkers for the detection of prenatal alcohol exposure: a review. *Alcohol Clin Exp Res* 2017;41(2):251-261 [doi: [10.1111/acer.13309](https://doi.org/10.1111/acer.13309)] [Medline: [28098942](https://pubmed.ncbi.nlm.nih.gov/28098942/)]
46. Tyrallis H, Papacharalampous G. Boosting algorithms in energy research: a systematic review. *Neural Comput & Applic* 2021;33(21):14101-14117 [doi: [10.1007/s00521-021-05995-8](https://doi.org/10.1007/s00521-021-05995-8)]
47. Tohka J, van Gils M. Evaluation of machine learning algorithms for health and wellness applications: a tutorial. *Comput Biol Med* 2021;132:104324 [FREE Full text] [doi: [10.1016/j.combiomed.2021.104324](https://doi.org/10.1016/j.combiomed.2021.104324)] [Medline: [33774270](https://pubmed.ncbi.nlm.nih.gov/33774270/)]
48. Accessing CIFASD Research Data. CIFASD. URL: <https://cifasd.org/data-sharing/> [accessed 2023-06-28]

Abbreviations

ASD: autism spectrum disorder

AUD: alcohol use disorder

AUPRC: area under the precision-recall curve

AUROC: receiver operating characteristics curve

CIFASD: Collaborative Initiative on Fetal Alcohol Spectrum Disorder

FAS: fetal alcohol syndrome

FASD: fetal alcohol spectrum disorder

GBM: gradient-boosting machine

IRB: institutional review board

ML: machine learning

PAE: prenatal alcohol exposure

Edited by T Leung; submitted 13.12.22; peer-reviewed by S Park, H Chen, B Kang; comments to author 17.04.23; revised version received 22.05.23; accepted 18.06.23; published 18.07.23

Please cite as:

Oh SS, Kuang I, Jeong H, Song JY, Ren B, Moon JY, Park EC, Kawachi I

Predicting Fetal Alcohol Spectrum Disorders Using Machine Learning Techniques: Multisite Retrospective Cohort Study
J Med Internet Res 2023;25:e45041

URL: <https://www.jmir.org/2023/1/e45041>

doi: [10.2196/45041](https://doi.org/10.2196/45041)

PMID:

©Sarah Soyeon Oh, Irene Kuang, Hyewon Jeong, Jin-Yeop Song, Boyu Ren, Jong Youn Moon, Eun-Cheol Park, Ichiro Kawachi. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 18.07.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.