Title

Machine learning approaches to the identification of children affected by prenatal alcohol exposure: A narrative review.

Journal

Alcoholism: Clinical and Experimental Research, 48(4)

Authors

Suttie, Michael

Kable, Julie

Mahnke, Amanda

et al.

Publication Date

2024-04-01

DOI

10.1111/acer.15271

Peer reviewed

HHS Public Access

Author manuscript

Alcohol Clin Exp Res (Hoboken). Author manuscript; available in PMC 2025 April 01.

Published in final edited form as:

Alcohol Clin Exp Res (Hoboken). 2024 April ; 48(4): 585–595. doi:10.1111/acer.15271.

Machine learning approaches in the identification of children affected by prenatal alcohol exposure: A narrative review

Michael Suttie, PhD[1,2], **Julie Kable, PhD**[3], **Amanda H. Mahnke, PhD**[4], **Gretchen Bandoli, PhD**[5]

[1]Nuffield Department of Women's & Reproductive Health, University of Oxford, UK

[2]Big Data Institute, University of Oxford, UK

[3]Departments of Psychiatry and Behavioral Science and Pediatrics, Emory University School of Medicine, 201 Dowman Drive, Atlanta, GA, 30322, USA

[4]Department of Neuroscience and Experimental Therapeutics, Texas A&M University School of Medicine, 8447 Riverside Parkway, Bryan, TX 77807, USA

[5]Department of Pediatrics, University of California San Diego, La Jolla, CA, USA

Abstract

Fetal alcohol spectrum disorders (FASDs) affect at least 0.8% of the population globally. The diagnosis of FASD is uniquely complex, with a heterogeneous physical and neurobehavioral presentation requiring multidisciplinary expertise for diagnosis. To expand early identification and diagnosis of FASD, many researchers have begun to incorporate machine learning approaches into FASD research to identify children with FASD or who are affected by prenatal alcohol exposure. This narrative review highlights these efforts. We first include an introduction to machine learning. We then summarize examples from the literature into neurobehavioral screening tools and physiologic markers of exposure. We discuss individual efforts, including models that classify FASD based on parent-reported neurocognitive or behavioral questionaries, 3D facial imaging, brain imaging, DNA methylation patterns, microRNA profiles, cardiac orienting response, and dysmorphic facial features. We highlight model performance and discuss the limitations of these approaches. We conclude with a broader consideration of the scalability of these approaches and considerations for how these machine learning models, largely developed from clinical samples or highly-exposed birth cohorts, may perform in the general population.

"Fetal alcohol spectrum disorders" (FASDs) is a collective term encompassing a range of diagnostic outcomes that result from prenatal alcohol exposure (PAE). FASDs affect approximately 8 of 1000 people in the global population, with estimates varying drastically by geographic location (Lange et al., 2017a; May et al., 2018) and by ascertainment method (Coles et al., 2022, 2016). One outcome within FASD is fetal alcohol syndrome (FAS), characterized by central nervous system anomalies, growth deficiency, neurobehavioral deficits, and characteristic facial dysmorphism (Hoyme et al., 2016; Jones, 2011). While FAS is identifiable by the presence of phenotypic traits, there is a significant variation

Corresponding author: Gretchen Bandoli, PhD, 9500 Gilman Drive, MC 0828, San Diego, CA 92093, gbandoli@ucsd.edu.

in the clinical presentation across FASDs, for which at least 11 classification systems are used for the diagnostic categorization. While inconsistencies between systems have been identified, most require confirmed prenatal exposure to alcohol, neurobehavioral impairment, growth impairment, and dysmorphic features (Coles et al., 2022, 2016). There is no consensus threshold for PAE, and the report can come directly from the biological mother, from a collateral source, or from medical records or other documented encounters suggestive of exposure, including justice system records of public intoxication or driving under the influence. Neurobehavioral impairment spans the domains of neurocognitive, behavioral, and adaptive function (Hyland et al., 2023). The presence of dysmorphic features specific to PAE aids in diagnosing individuals presenting with neurobehavioral impairments. The three essential elements of dysmorphology include cardinal features (short palpebral fissures, smooth philtrum, and thin upper lip vermilion), brain size or structure (or head circumference), and growth deficits (Hoyme et al., 2016).

Due to the reliance on these three key diagnostic components, diagnosing individuals affected by PAE is a complex medical process requiring expertise across disciplines. There are additional difficulties in obtaining an accurate diagnosis of FASDs, including lack of awareness or training on FASDs among clinicians, subtlety and variability of the physical features, particularly as children age, lack of available data or under-reporting of prenatal alcohol exposure, and non-specific or variable expression of the alcohol-related neurobehavioral and growth abnormalities. Consequently, FASDs are grossly under-recognized (Chasnoff et al., 2015). In a study by Chasnoff and colleagues, 547 children in foster care were evaluated for FASDs. From the sample, 156 met the criteria for FASD, of which 125 (80%) had never been diagnosed with a FASD (Chasnoff et al., 2015). Early diagnosis is the most important predictor of optimal treatment outcomes, and may further reduce subsequent adverse outcome across the life course (Olson and Montague, 2011; Streissguth et al., 2004). Therefore, promoting processes that facilitate identifying and diagnosing children with FASDs is critical.

With the goal of expanding early identification and diagnosis of FASDs, many researchers have begun to incorporate machine learning approaches into FASD research. Others have performed systematic or scoping reviews of these efforts (Kable and Jones, 2023; Roomaney et al., 2022). Here, we present a narrative review of the literature. To enhance the accessibility of these studies to researchers without a data science background, we first introduce machine learning, particularly as it relates to predictive models. We then will highlight select examples of studies in the field. Although heterogenous and not mutually exclusive, we group the efforts behind predictive models for FASDs or PAE into two categories: 1) neurobehavioral screening tools, and 2) physiologic markers of exposure. The performance and limitations of each approach are highlighted. We will then discuss scalability of these approaches and considerations for how these machine learning models may perform in the general population, and future directions for the field.

Machine learning

Although machine learning shares underlying statistics with classic statistics, the primary difference is that classical statistical models are used for inference, while machine learning

can be used for prediction (Bzdok et al., 2018). Moreover, statistical models fit project-specific probability models that often collapse as the numbers of input variables and the associations between them increase. In contrast, machine learning (also referred to as statistical learning), makes minimal assumptions about the data-generating system and can be effective in the presence of complicated interactions often underlying human health outcomes (Bzdok et al., 2018). Further, machine learning typically prioritizes predictive accuracy over causal inference, often exploiting variables one may consider confounders or mediators in inferential statistics (Bi et al., 2019).

Generally, machine learning algorithms can be divided into supervised and unsupervised learning. In unsupervised models, data are not labeled; rather, the algorithm searches for commonalities or patterns within the data, agnostic to an associated outcome or exposure. Conversely, in supervised models, labeled data are provided to train models (e.g. the outcome is known and has a specified value (e.g. 0 or 1)). One of the primary applications of supervised models in biomedical sciences is to create classifiers that can separate subjects into two or more classes based on attributes (or features) measured in each subject (Foster et al., 2014) (Figure 1). This results from the machine ‘learning’ important features of a dataset to enable it to make predictions about other data that were not included in the training data set. Machine learning uses various statistical techniques that allow the computer to derive the algorithm that most efficiently identifies a group or solves a given predictive problem. Algorithms in classification models range from logistic regression, naïve Bayes, decision trees and support vector machines to more complex architectures using neural networks and ensemble models that combine base estimators (e.g. random forest; Table 1).

Classifier model accuracy is evaluated through various metrics. Many report the true positives, false positives, true negatives and false negatives, which can be shown in a confusion matrix. For binary classification, receiver operating curves plot the true positive rate versus the false positive rate with accuracy quantified as the area under the curve (AUC), where a value of 0.5 indicates the model performs no better than random chance, 1.0 indicating perfect classification, 0.7–0.8 indicating acceptable performance, and ≥0.8 indicating excellent or outstanding performance (Hosmer Jr et al., 2013). Alternatively, in the case of class imbalances where the group status is not evenly balanced, the precision-recall curve is used to mitigate bias in the trained model (Lever et al., 2016). The three main metrics reported are accuracy (the percentage of correct predictions for the test data), precision (the fraction of true positives among all predicted positives or the positive predictive value), and recall (true positives among all actual positives, or the sensitivity). Sensitivity (the proportion with the outcome correctly identified) and specificity (the proportion without the outcome correctly identified) demonstrate the ability of the model to correctly label individuals who do or do not have the outcome.

One vulnerability of algorithms, particularly with highly dimensional data, is overfitting the data, where the algorithm performs *too* well on the training data by fitting the noise of the dataset and, conversely, performs poorly on unseen data. Approaches to minimize this are splitting the data into training and test subsets (where classifiers are refined in the training data, and then evaluated on unseen, ‘test’ data), using cross-validation (a resampling method where different portions of the data are used to train or test the data on different iterations),

and validating the algorithm on external data from a completely different sample. Studies should, at a minimum, test the algorithm in a hold-out sample, and ideally, test it in a completely different sample to truly determine the accuracy and application of the model.

One issue often encountered in biomedical models is that of class imbalance- where the outcome occurs with much greater or less frequency than 50%. Classes that make up a larger proportion are the majority class, while those that make up the smaller portion are the minority class. In highly skewed data, classifiers can have a very high accuracy from always predicting the majority class. Ways to resolve an imbalance include random undersampling of the majority class or oversampling the minority class. Undersampling causes a loss in information, and often researchers are working with small datasets and cannot afford to further reduce the sample, particularly when splitting data further into training and test data. Oversampling the minority class is an approach that often involves the creation of synthetic samples through small variations in the observed data. Examples of this include synthetic minority oversampling technique (SMOTE), which creates new cases of the minority class in the training data using a nearest neighbor approach to balance the dataset. Since its publication in 2003, more than 100 variants of SMOTE have been developed (Kovács, 2019).

Neurobehavioral screening

Although FAS can be diagnosed from dysmorphic features at birth, other diagnoses on the spectrum require evidence of neurocognitive or neurobehavioral impairments. While severe impairments can be assessed early in life using standardized measures of infant development, less severe impairments cannot be assessed using standard instruments until later in childhood (Gomez and Abdul-Rahman, 2021), contributing to delayed diagnosis. The diagnosis of alcohol-related neurodevelopmental disorder (ARND), which represents 80–90% of FASD cases (Chudley, 2008), is based primarily on neurodevelopmental impairments as the dysmorphic features are typically absent (Lange et al., 2017b). Three functional domains for impairment are neurocognition, self-regulation and adaptive functioning (Kable and Coles, 2018). However, many other diagnoses share these impairments, including ADHD, oppositional defiant disorder, and conduct disorder. The difficulty in obtaining confirmation of PAE results in the misdiagnosis of FASDs with these and other neurodevelopmental disorders (Lange et al., 2017b). Further, neurodevelopmental assessment is costly in time and money, and a shortage of clinicians to perform the evaluations further reduces the likelihood of a child with FASD receiving a diagnosis (Lange et al., 2017b; Petrenko et al., 2014).

Due to these barriers in the healthcare system, substantial effort has been focused on creating predictive models that can accurately discriminate children affected by PAE, often compared to typically developing unexposed children and children with ADHD. In 2006, Nash and colleagues analyzed a sample of children aged 6–16 with FAS or ARND (n=30), ADHD (n=30), and typically developing controls (n=30). Using discriminant function analysis and AUC, they analyzed the parent-completed Child Behavior Checklist (CBCL) and found that seven items differentiated children with FASD from controls (sensitivity= 86%, specificity=82%). Six items differentiating FASDs from ADHD (with

some overlap of items that differentiated FASD from controls) had slightly worse performance (sensitivity=70%; specificity=80%) (Nash et al., 2006). These same items from the CBCL had high sensitivity and specificity for discriminating FASD compared to typically developing children when later applied to a sample of children ages 4–6 (94% sensitivity and 96% specificity) (Breiner et al., 2013). When the same items were validated in a sample of children with FASDs, PAE without FASDs, and typically developing controls, sensitivity was reduced for children with FASDs (62.5%) and PAE without FASDs (50%), although specificity was high for typically developing children (100%) (LaFrance et al., 2014). Around the same time, Mattson and colleagues performed latent profile analysis on 547 neuropsychologic variables in a sample of children with and without FAS (Mattson et al., 2010). Using logistic regression, a 2-class model successfully distinguished FAS from nonexposed controls with 92% accuracy, with 87.8% of FAS cases and 95.7% of controls correctly classified. In a second analysis, the same profile distinguished children with PAE but without FAS from non-exposed controls (84.7% accuracy). The authors noted the sample was not large enough to validate the model in a hold-out sample (Mattson et al., 2010).

Other neurobehavioral screening measures have been combined with physical features to distinguish children affected by PAE from those who are not. Mattson and colleagues tested the inclusion of more than 1,000 composite and subtest scores from measures that assessed behavior, cognition, and dysmorphology in the development of the decision tree (Goh et al., 2016). The final model contained only four measures - the CBCL, Vineland Adaptive Behavior Scales, and IQ score and dysmorphology. In validation samples, sensitivity was 64–81%, and specificity was 78–80%. Young age and co-morbid ADHD contributed to misclassification (Goh et al., 2016). The model (named the FASD Tree) was recently validated again in a sample of children with (n=224; 186 with FASD diagnosis) and without (n=78) PAE (Mattson et al., 2023). Sensitivity (78.1%) and specificity (70.5%) remained similar to the original work for classifying PAE. In addition, the FASD Tree had high sensitivity (82.8%) and good specificity (62.3%) when classifying FASDs. Recently, the features from the FASD Tree were also used to develop a risk score to distinguish alcohol-exposed children from those who were not (Bernes et al., 2022). In a validation sample, setting the score threshold for defining risk or deviance, also known as a cut point, at 1.5 points had an accuracy of 76.6%, sensitivity of 76.9%, and specificity of 76.5%. When the cut point was increased to 2.5, sensitivity fell (63.6%) while specificity increased (87.7%). The score also correlated with IQ and executive functioning scores (Bernes et al., 2022).

Similar efforts were undertaken by researchers in Germany, who developed a machine learning algorithm (FASDetect) for the detection of FASDs in patients ages 0–19 with ADHD symptoms (Ehrig et al., 2023). Researchers used a clinical sample of 275 children with FASDs with or without ADHD, and 170 children with ADHD without FASDs. Six machine learning algorithms were tested (logistic regression, SVM, random forest, gradient boosting decision tree, k-nearest neighbor classification, and Gaussian process classification algorithms). The random forest model performed best, and the final model had six features- body length and head circumference at birth, IQ less than 85 points, socially intrusive behavior, poor memory, and sleep disturbance- and was sufficient to differentiate youth with versus without FASDs. The cross-validated AUC was 0.93 (95% CI 0.85, 1.00). Although 10-fold cross-validation was employed, there was no hold-out data or external validation,

which are important measures of model performance and generalizability (Moons et al., 2012a).

These studies represent some of the efforts that have been made to develop neurobehavioral screening tools that can distinguish children with FASDs from typically developing children or children with other neurodevelopmental disorders. The lack of reliance on information about PAE in these screeners is important, as reliable information can often be a barrier to identifying alcohol-affected children. Some of the screeners have included external validation samples, which are crucial to ensuring the model is not overfitting the data and can be used in other populations potentially affected by PAE. However, to date, the majority have focused on clinical samples with a disproportionate number of alcohol-affected children. These screeners are designed to work in clinical offices; however, it is unclear how they will behave in a general population where FASDs are relatively rare, and children may have lesser impairments than those captured in clinical samples.

Physiologic markers

Although FASD diagnosis has commonly relied on self-report of PAE in the absence of cardinal facial features, this self-report is often difficult to obtain due to stigma, out-of-home placements of the affected child, and long recall periods. PAE affects many systems of the developing fetus, including the central nervous system, introducing the possibility of using physiologic markers of PAE as a substitute for self-report. Historically, matrices like meconium, urine, or blood were used to detect products of ethanol metabolism (Bakhireva and Savage, 2011; Concheiro-Guisan and Concheiro, 2014; Lussier et al., 2018). However, they were limited by short half-life or detection of exposure only proximate to parturition, which may not capture early exposure resulting in FASDs. Physiologic markers have expanded rapidly over the past two decades, and now include markers thought to be more enduring or more sensitive to detecting children affected by alcohol. These include the use of 3D facial or brain imaging and epigenetic markers such as DNA methylation or microRNAs.

1. Facial and brain imaging

PAE results in craniofacial anomalies that can be used to discriminate exposed children. As early as 2007, researchers began using 3D images, allowing precise depth measures that were impossible in 2D imaging alone (Roomaney et al., 2022). Early examples include a study in 149 Cape Coloured (mixed ancestry) children from South Africa (86 FAS and 63 control) where a unique set of facial regions and features accurately discriminated FAS and control faces without any human intervention (Fang et al., 2008). Linear measurements were computed from anatomical landmarks derived from manually annotated 3D images, providing a series of morphometric parameters to three different machine learning approaches to evaluate support vector machine (SVM), k-nearest neighbor, and decision tree approaches. The classifier model had a sensitivity of 82.7% and a sensitivity of 76.2% in a hold-out sample. A subsequent study on a similar South African dataset utilized whole surface modeling from 3D images rather than a landmark-based approach to assessing facial dysmorphism across FASDs (Suttie et al., 2013). Dimensionality reduction

plays a vital role in data analysis and machine learning, and this study utilized a principal component analysis (PCA) based method known as dense surface modeling (DSM) to compute a series of principal components that represent facial shape. The DSM algorithm builds surface models from raw 3D data, initially aligning and warping surfaces guided by a series of manually placed anthropometric landmarks. The result is a dense correspondence of points, matching points on different face surfaces to produce a shape-based PCA of variation of point displacement from the mean face. Suttie and colleagues utilized DSM of the face on 192 participants to perform control-FAS discrimination testing of face shape using 20 randomly sampled 90% to 10% training to unseen test subsets. The whole face and sub-regions analyzed with discrimination testing using SVM were predicted to be near perfect, with the peri-orbit (eye region) achieving AUC > 0.98. This study also utilized unsupervised learning methods, referred to as facial signature graphs, where clusters of individuals are grouped based on their expression of facial dysmorphism and proximity to one another. This method uncovered the more subtle dysmorphism that arises from PAE across the FASD spectrum, and not just limited to FAS (Suttie et al., 2013). More recent work using DSM techniques included 166 children from South Africa and 249 children of European ancestry (Suttie et al., 2017), comparing ethnic differences between cohorts. Classification algorithms were created to test the control FAS agreement of three different algorithms with AUC >0.95 for individuals of South African or European ancestry. Separately, Blanck-Lubarsch and colleagues compared decision trees, SVM, and k-nearest neighbors for both accuracy and clinical applicability in analyzing 3D facial scans in FAS cases (n=30) and controls (n=30) (Blanck-Lubarsch et al., 2021). All three methods were found to have accuracy above 89.5% in a hold-out sample; decision trees were found to be more practical to use clinically, as they provide an easily implemented, simplistic approach making it amenable for frontline clinicians (Blanck-Lubarsch et al., 2021).

More sophisticated machine learning approaches to 3D facial analysis have recently adopted deep neural networks (DNNs) for dimensionality reduction using the auto-encoder approach (Kingma and Welling, 2019). This DNN architecture typically comprises both an encoder and a decoder. The encoder’s primary function is to condense the complex 3D facial shape data into lower-dimensional representations of facial morphology. Subsequently, the decoder is responsible for restoring these representations, thereby reconstructing the original 3D facial shape. Lui and colleagues employed this technique to a dataset containing 9-year-old (n=3149), and 13-year-old children (n=2477), with 1878 children assessed at both ages (Liu et al., 2023). The resulting auto-encoder dimensionality reduction provided 200 facial traits representing each 3D face. They performed an independent linear regression on these traits to assess the impact of PAE on facial morphology. No significant associations were found in the 13-year-old children, but at the 9 years, PAE was significantly associated with several facial traits, even at low levels. Most notably, they discovered an association between facial traits and alcohol use only in the 3 months before pregnancy.

In addition to differences in facial morphology, PAE also affects the developing brain. Small head circumference is one of the features of FASDs. Given the correlation between head circumference and brain volume in young children (Bartholomeusz et al., 2002), researchers have evaluated whether regional brain volume can be used in classifier models of FASDs. In a sample of 160 children (79 FASDs, 81 controls) from Canada with a separate validation

set (67 FASDs, 74 controls), a binary classification model based on brain volumes was created to discriminate between typically developing individuals and those with FASDs (Little and Beaulieu, 2020). In these models, all FASD diagnoses were grouped together, although about 25% of training cases had sentinel features. The model's ten most heavily weighted brain regions included three subcortical gray matter regions, three cortical gray matter regions located in the temporal lobe, two cortical regions located in the frontal lobe, and two regions along the cingulate gyrus. The model had moderate performance on the independent test data (accuracy=77%, sensitivity=64%, and specificity=88%).

Researchers have begun combining these promising results from facial and brain models. A study by Suttie and colleagues (Suttie et al., 2018) assessed face and brain morphology separately and as a multi-modal representation. Localized regions of the face were combined using DSM with 3D representations of the corpus callosum and caudate nucleus. In each case, the combined face-brain models had better discrimination of children with FAS than the single face and brain representations alone. Results indicated midline facial differences were correlative with midline defects of the brain, and the most significant improvements on single model representations were made when midline facial representations were combined, i.e, the corpus callosum combined with, the nose (AUC=1.00), lip-vermillion (AUC>0.90), and philtrum (AUC=0.98).

Notably, most of these approaches aimed to discriminate FAS (which has the presence of cardinal facial features in the diagnosis) from controls. However, FAS occurs the least frequently among FASDs, potentially limiting the application of 3D facial screening models to the larger FASD population. However, a new proprietary software tool called Face2Gene (FDNA Inc, Boston, MA) combines facial recognition from 2D photographs to evaluate the presence of dysmorphic features. Researchers recently used this approach to test discrimination of all FASD diagnoses (FAS (n=36), partial (p)FAS (n=31), and alcohol-related neurodevelopmental disorder (ARND; n=22)) against controls (n=50), all 5–9 years of age (Valentine et al., 2017). It should be noted that in this study, the definition of ARND included physical symptoms not typically included, such as small occipital frontal head circumference. The comparisons were of the dysmorphology scoring system, either manually evaluated by trained dysmorphologists, or the computer-aided by combining the Face2Gene results for facial images with dysmorphologist evaluated non-facial features. Sensitivity was greatest for FAS or pFAS (78–79%) vs controls, and lowest for ARND (50%). Specificity was high (78–92%) for all. Importantly, for FAS and pFAS, the software was nearly as accurate as diagnosis made by expert dysmorphologists, and for ARND, it was slightly better (Valentine et al., 2017).

2. Epigenomic markers

The epigenome is the collection of modifications and modifiers that regulate DNA expression without altering the DNA sequence and include DNA methylation and microRNAs (miRNAs). Research in clinical and preclinical models has suggested that PAE can result in a potentially life-long DNA methylation signature in the central nervous system and peripheral tissues (Lussier et al., 2017). A potential signature of DNA methylation was identified in participants from the NeuroDevNet (now Kids Brain Health Network)

Canadian FASD study (Portales-Casamar et al., 2016). Using buccal endothelial cells from 110 children with FASDs and 96 sex and aged-matched controls, genome-wide DNA methylation analysis identified 658 CpG sites that were differentially methylated in the children with FASDs compared to the controls. In a follow-up study, Lussier and colleagues assessed buccal swabs from 24 children with FASDs and 24 age and sex-matched controls (Lussier et al., 2018) as a validation cohort. A signature containing 648 of the 658 previously identified CpG loci identified children with FASDs with a sensitivity of 91.7%, specificity of 75%, and AUC of 92%. They further tested the accuracy of their methylation signature for FASDs classification using data from a cohort of individuals with autism spectrum disorders (ASDs) and typically-developing controls. Only one individual with ASD was incorrectly classified as having a FASD (specificity=99%). The authors did not detect any bias from sex, age or ethnicity and concluded the findings supported a distinct methylation pattern for children with FASDs (Lussier et al., 2018).

To date, a few human studies have evaluated miRNAs as biomarkers of PAE (Balaraman et al., 2016; Gardiner et al., 2016; Mahnke et al., 2021). miRNAs are a class of small non-protein-coding RNAs that intracellularly act as repressors of protein translation but are also released from cells into circulation where they are thought to act as endocrine factors. Potential miRNA biomarkers of the effects of PAE were identified in maternal plasma from 68 mothers from a longitudinal cohort in Ukraine (22 mothers with heavy alcohol exposure and alcohol-affected child (HEa), 23 mothers with heavy alcohol exposure and apparently unaffected children (HEua), and 23 mothers with low or no alcohol exposure and unaffected children (UE)) (Balaraman et al., 2016). Random forest models were used to identify a combination of miRNA expression at mid or late pregnancy and clinical variables, e.g. maternal smoking, socioeconomic status, fetal sex, that predicted future infant outcomes. For this random forest analysis, the model was trained using subgroups created by subsampling with replacement and aggregated across many subgroups, or bagging. Model performance was determined by a subsample that was not used in the training, known as the out-of-bag sample. In the out-of-bag sample, the HEa and UE groups were classified into their respective groups with an overall misclassification rate (proportion of misclassified observations) of 13.3%. The classification model more accurately assigned membership of UE samples to the UE group, with a classification error rate of 8.7%, whereas the error rate for the HEa group was 18.2%. The identified predictive variables were also applied to the HEua group and stratified the HEua group into four subgroups: UE-like, HEa-like at mid-pregnancy resolving to UE-like at late pregnancy, UE like at mid-pregnancy resolving to HEa-like at late pregnancy, and HEa-like. This subcategorization of the HEua group suggests that a profile of miRNAs and clinical variables may be able to identify apparently unaffected infants that may not be diagnosed for FAS in infancy but are still at risk for negative outcomes. The authors concluded that maternal plasma miRNAs predicted infant outcomes and may be useful in classifying difficult-to-diagnose FASD subpopulations (Balaraman et al., 2016).

Other markers

While by no means exhaustive, others have taken different approaches to assessing markers of exposure. Kable and colleagues have been working to identify markers of PAE that

could be scalable across populations. Such work includes the assessment of the predictive validity of the cardiac orienting response (COR). COR has been shown previously to be sensitive to PAE (Sokolov, 2002). They are based on electrocardiogram recordings during the presentation of auditory or visual stimuli and are relatively easy to collect on a large scale and in low-resource settings (Mesa et al., 2017). In a sample of 124 infants from a birth cohort in Ukraine, the COR collected at 6–12 months of age exhibited an AUC score of 0.81, negative predictive value of 85%, and positive predictive value of 66% with cross-fold validation when classifying neurodevelopmental delay (measured with the Bayley Scales of Infant Development). In this analysis, SMOTE was used to balance the dataset, and a weighted logistic regression was employed as the algorithm. Adding in indices of maternal drinking did little to improve performance (Mesa et al., 2017). In a second study, the authors repeated classifier models to determine if the COR collected at 6 and 12 months predicted FASD diagnosis at 3–4 years of age. There, ROC analysis of the visual response yielded an AUC value of 0.77 for predicting to pFAS/FAS status (Kable et al., 2021).

Dysmorphic features and the dysmorphology score have been used to predict FASDs. These efforts reflect both the presence of dysmorphic features in the FASD diagnosis, and research demonstrating that the presence of cardinal features were associated with poorer neurodevelopmental functioning (Chasnoff et al., 2010). Kalberg and colleagues studied whether the dysmorphology score (Hoyme et al., 2016) predicted FASD diagnosis. The score is a linear combination of the dysmorphic features with assigned weights (ranging from 1–3) (Kalberg et al., 2019). From a cohort of children followed from birth through age 5 in South Africa (n=155; FASD=79), they found that the dysmorphology score at 9 months of age predicted FASD diagnosis at age 5 with an AUC of 0.78 (Kalberg et al., 2019). The FASD sample was predominantly children with FAS (n=34) and pFAS (n=13), and by 18 months, the dysmorphology score differentiated the children with FAS or pFAS both from controls and from children with ARND (n=18). It does not appear that validation was done in hold-out samples or an external sample (Kalberg et al., 2019). In a similar study, Bandoli and colleagues sought to determine whether the full list of alcohol-related dysmorphic features in infancy, with or without information on PAE, correctly classified children ages 3–4 years on neurodevelopmental outcomes and FASDs (Bandoli et al., 2022). Using the birth cohort from Ukraine (n=273; FASD=62) and a logistic regression classifier, sensitivity ranged from 12–19% for the full model predicting neurodevelopmental delay, and 33–63% when assessed in children with high PAE. The models were minimally altered by removing PAE information and only relying on growth and dysmorphic features. Finally, sensitivity for discriminating FASD was 27% in the full sample and 62% when limited to children with high PAE (Bandoli et al., 2022). In this approach, the classifier was not examined in test or validation datasets. In 2023, Bandoli and colleagues expanded these features to include a broad range of pregnancy and infancy characteristics to predict FASD in preschool aged children (Bandoli et al., 2023). Data were split into separate training (80%: n=245) and test (20%: n=58; 11 FASD, 47 no FASD) datasets. Training data was balanced using data augmentation through SMOTE. Four classifier models (random forest, extreme gradient boosting (XGBoost), logistic regression (full model) and backwards stepwise logistic regression) were evaluated for accuracy, sensitivity, and specificity in the hold-out sample. Random forest models had the highest sensitivity (0.54), with accuracy of 0.86

(95% CI 0.74, 0.94) in hold-out data. The best performing algorithm correctly classified 6 of 11 children with FASD, and 44 of 47 children without FASD. The variables contributing the most to the model included measures of PAE and infant neurodevelopmental assessment scores. Although sensitivity in the full sample (54%) was improved compared to 27% observed with only dysmorphic features in earlier work (Bandoli et al., 2022), almost half of children who would later receive a diagnosis were not correctly identified (Bandoli et al., 2023).

FASD is a multifaceted diagnosis reliant on multiple domains. Multi-modal analysis is vital to ensure high sensitivity and specificity, as identifying features from a single domain is insufficient for providing a full insight into a patient's presentation. To date, most classifier models have focused on single domains, like facial imaging, neurodevelopmental screener performance, DNA methylation, etc. Given the relative success of these individual approaches, combining models appears a reasonable approach to continuing to improve the performance of the models. This was the approach of researchers using data from children ages 5–18 recruited in Canada, where they combined data from eye movement behaviors, psychometric test scores, and diffusion tensor imaging (DTI) of the brain to construct a multimodal classifier (Zhang et al., 2019). Researchers prioritized large-scale applicability, using tools from the theory of value of information to evaluate the cost-benefit metrics of the approach. Data from saccadic eye movements, DTI and psychometric tests were analyzed using SVM-recursive feature elimination. Classification accuracy ranged from 65% (saccadic eye movement), 67% (DTI) to 78% (psychometric data). The full multimodal model (combining all assessments) was assessed from the probability of the participant being identified as having an FASD predicted from each assessment and concatenated as input for training a logistic regression classifier. The final classifier had accuracy of 84.8% (sensitivity 81.8%, specificity 87.5%) in test data. However, due to the cost of many paradigms included, researchers proposed using only saccadic eye movement and natural viewing tasks (sensitivity=77%, specificity=79%), with children screened as 'high risk' being recommended for a full evaluation. For older children, authors suggested psychometric batteries could be added in to the saccadic eye movement and natural viewing tasks for the initial screening (Zhang et al., 2019).

Current limitations and future directions

For at least two decades, researchers have used machine learning to develop classifier models to better screen or identify children with FASDs. Most models appear to have modest to good discrimination of children affected by PAE or with FASDs. A few additional considerations remain when interpreting the efforts to date- cost vs. scalability and translation into general populations.

As with all successful screening tools, balancing the accuracy and the cost of the approach remains a challenge. The Zhang et al. study highlighted the problem that many domains have moderate to good predictive ability but come at very high costs (Zhang et al., 2019). MRI scans and 3D imaging are not widely available, particularly in low-resource settings. Likewise, biomarkers like DNA methylation and miRNA are not yet scalable. And models that rely on expert dysmorphologists also severely limit the broad application of

the screener. Further, for model development in training data, large samples containing the measure are required. Efforts like the Collaborative Initiative on Fetal Alcohol Spectrum Disorders (CIFASD), which was created in 2003 to advance the understanding of FASDs, is largely responsible for much of the work reviewed here. Yet even with these efforts, challenges remain in finding datasets with labeled PAE or FASDs that also contain low-cost measures to exploit for model development.

In addition to identifying screeners that can be widely scaled across high and low resource settings, the current work is further limited by a lack of translation for general populations. Most studies described in this review were conducted in birth cohorts enrolled based on heavy prenatal alcohol use, or from clinical settings where children were being assessed due to cognitive or behavioral concerns. Although these classifiers work well in those settings, they cannot be expected to have the same performance in a general population. Further, while a few of the studies were noted as having external validation samples, most did not. Machine learning algorithms suffer from over-learning noise in a dataset. While splitting samples into training-test subsets or using cross-validation can minimize that risk, validation in an independent dataset, preferably from a different setting or population, is required to fully evaluate the broad performance of the model. Readers looking to adopt these practices can learn more about them (Moons et al., 2012a, 2012b; Poldrack et al., 2020).

Researchers in this space may benefit from reviewing the ASD and ADHD machine learning literature, which has similar efforts towards screening and identifying affected children. There, machine learning algorithms have shown promise, including the ability to differentiate autism both from typically developing children (Kosmicki et al., 2015; Wall et al., 2012a, 2012b) as well as from children with ADHD (Duda et al., 2016). It has also been used to develop multivariable profiles of children with ADHD (Nilsson, 2005; Yasumura et al., 2017) and to predict autism spectrum disorder (ASD) based on facial features (Ahmed et al., 2022). Like FASD, the ASD literature has multifaceted efforts, including a large literature on classifier models and neuroimaging (Song et al., 2021), eye-tracking (Wei et al., 2023) and behavioral inputs. The latter was discussed in a recent systematic review, which included 22 studies (Cavus et al., 2021). The authors noted that despite many promising results from neuroimaging, eye tracking, and genetic data, that behavioral data also showed promise particularly given the scalability. However, like FASD research to date, authors noted that while individual results of ASD machine learning studies have been promising, none have demonstrated broad clinical relevance to date (Cavus et al., 2021). Similar findings have been noted in ADHD machine learning research. In a recent review of 92 studies (the majority MRI, followed by physiological signals and questionnaire data), authors noted that most focused on single modality studies and had accuracy of 80–90%, yet lacked broad clinical adoption (Loh et al., 2022). As the respective fields continue efforts towards developing scalable, clinically useful screening tools, they should look to each other for new tools or ideas that could be beneficial to efforts across the diagnoses.

In summary, many efforts have been made to create FASD screeners, primarily based on neurodevelopmental profiles or physiologic markers. Many of the classifiers have shown promise, however, challenges remain in developing models that are both accurate and cost effective. Given the heterogeneity of FASD presentation, and the multi-faceted

nature of the diagnosis, a prominent direction for machine learning in this field should be combining accessible modalities such as neurodevelopmental assessment and facial imaging to improve sensitivity and specificity. Ultimately, FASDs remain underdiagnosed, and countless children are not afforded interventions that can improve their quality of life. Continued collaborations between researchers, clinicians, and among groups like CIFASD are critical to advancing the field.

Acknowledgement

Gretchen Bandoli is funded by the National Institutes on Alcohol Abuse and Alcoholism (NIAAA) (1 K01 AA027811-01). Amanda Mahnke is funded by Texas A&M University's Accountability, Climate, Equity, and Scholarship (ACES) Faculty Fellows Program and start-up funding. Part of this work was done in conjunction with the Collaborative Initiative on Fetal Alcohol Spectrum Disorders (CIFASD), which is funded by grants from the NIAAA. This work was supported by CIFASD NIAAA grants U01AA014809 (Michael Suttie) and U01AA026108 (Julie Kable). We wish to thank Elizabeth Powell for her contributions to the content and editing of this manuscript.

References

Ahmed ZAT, Aldhyani THH, Jadhav ME, Alzahrani MY, Alzahrani ME, Althobaiti MM, Alassery F, Alshaflut A, Alzahrani NM, Al-Madani AM (2022) Facial Features Detection System To Identify Children With Autism Spectrum Disorder: Deep Learning Models. Comput Math Methods Med 2022:3941049. [PubMed: 35419082]

Bakhireva LN, Savage DD (2011) Focus on: biomarkers of fetal alcohol exposure and fetal alcohol effects. Alcohol Res Heal J Natl Inst Alcohol Abus Alcohol 34:56–63.

Balaraman S, Schafer JJ, Tseng AM, Wertelecki W, Yevtushok L, Zymak-Zakutnya N, Chambers CD, Miranda RC (2016) Plasma miRNA Profiles in Pregnant Women Predict Infant Outcomes following Prenatal Alcohol Exposure. PLoS One 11:e0165081. [PubMed: 27828986]

Bandoli G, Coles C, Kable J, Jones KL, Delker E, Wertelecki W, Yevtushok L, Zymak-Zakutnya N, Granovska I, Plotka L, Chambers C (2022) Alcohol-related dysmorphic features as predictors of neurodevelopmental delay in infants and preschool-aged children: Results from a birth cohort in Ukraine. Alcohol Clin Exp Res 46:2236–2244. [PubMed: 36308058]

Bandoli G, Coles C, Kable J, Lyons Jones K, Wertelecki W, Yevtushok L, Zymak-Zakutnya N, Granovska I, Plotka L, Chambers C, CIFASD T (2023) Predicting fetal alcohol spectrum disorders in preschool-aged children from early life factors. Alcohol Clin Exp Res n/a.

Bartholomeusz HH, Courchesne E, Karns CM (2002) Relationship between head circumference and brain volume in healthy normal toddlers, children, and adults. Neuropediatrics 33:239–241. [PubMed: 12536365]

Bernes GA, Courchesne-Krak NS, Hyland MT, Villodas MT, Coles CD, Kable JA, May PA, Kalberg WO, Sowell ER, Wozniak JR, Jones KL, Riley EP, Mattson SN (2022) Development and validation of a postnatal risk score that identifies children with prenatal alcohol exposure. Alcohol Clin Exp Res 46:52–65. [PubMed: 34806190]

Bi Q, Goodman KE, Kaminsky J, Lessler J (2019) What is Machine Learning? A Primer for the Epidemiologist. Am J Epidemiol 188:2222–2239. [PubMed: 31509183]

Blanck-Lubarsch M, Dirksen D, Feldmann R, Bormann E, Hohoff A (2021) Simplifying Diagnosis of Fetal Alcohol Syndrome Using Machine Learning Methods. Front Pediatr 9:707566. [PubMed: 35127583]

Breiner P, Nulman I, Koren G (2013) Identifying the neurobehavioral phenotype of fetal alcohol spectrum disorder in young children. J Popul Ther Clin Pharmacol = J la Ther des Popul la Pharmacol Clin 20:e334–9.

Bzdok D, Altman N, Krzywinski M (2018) Points of Significance: Statistics versus machine learning. Nat Methods 15:233–234. [PubMed: 30100822]

Cavus N, Lawan AA, Ibrahim Z, Dahiru A, Tahir S, Abdulrazak UI, Hussaini A (2021) A Systematic Literature Review on the Application of Machine-Learning Models in Behavioral Assessment of Autism Spectrum Disorder. J Pers Med 11.

Chasnoff IJ, Wells AM, King L (2015) Misdiagnosis and Missed Diagnoses in Foster and Adopted Children With Prenatal Alcohol Exposure. Pediatrics 135:264–270. [PubMed: 25583914]

Chasnoff IJ, Wells AM, Telford E, Schmidt C, Messer G (2010) Neurodevelopmental functioning in children with FAS, pFAS, and ARND. J Dev Behav Pediatr 31:192–201. [PubMed: 20375733]

Chudley AE (2008) Fetal alcohol spectrum disorder: counting the invisible - mission impossible? Arch Dis Child 93:721–722. [PubMed: 18719155]

Coles CD, Bandoli G, Kable JA, Del Campo M, Suttie M, Chambers CD (2022) Comparison of three systems for the diagnosis of fetal alcohol spectrum disorders in a community sample. Alcohol Clin Exp Res

Coles CD, Gailey AR, Mulle JG, Kable JA, Lynch ME, Jones KL (2016) A Comparison Among 5 Methods for the Clinical Diagnosis of Fetal Alcohol Spectrum Disorders. Alcohol Clin Exp Res 40:1000–1009. [PubMed: 27028727]

Concheiro-Guisan A, Concheiro M (2014) Bioanalysis during pregnancy: recent advances and novel sampling strategies. Bioanalysis 6:3133–3153. [PubMed: 25529882]

Duda M, Ma R, Haber N, Wall DP (2016) Use of machine learning for behavioral distinction of autism and ADHD. Transl Psychiatry 6:e732. [PubMed: 26859815]

Ehrig L, Wagner A-C, Wolter H, Correll CU, Geisel O, Konigorski S (2023) FASDetect as a machine learning-based screening app for FASD in youth with ADHD. NPJ Digit Med 6:130. [PubMed: 37468605]

Fang S, McLaughlin J, Fang J, Huang J, Autti-Rämö I, Fagerlund A, Jacobson SW, Robinson LK, Hoyme HE, Mattson SN, Riley E, Zhou F, Ward R, Moore ES, Foroud T (2008) Automated diagnosis of fetal alcohol syndrome using 3D facial image analysis. Orthod Craniofac Res 11:162–171. [PubMed: 18713153]

Foster KR, Koprowski R, Skufca JD (2014) Machine learning, medical diagnosis, and biomedical engineering research - commentary. Biomed Eng Online 13:1–9. [PubMed: 24410918]

Gardiner AS, Gutierrez HL, Luo L, Davies S, Savage DD, Bakhireva LN, Perrone-Bizzozero NI (2016) Alcohol Use During Pregnancy is Associated with Specific Alterations in MicroRNA Levels in Maternal Serum. Alcohol Clin Exp Res 40:826–837. [PubMed: 27038596]

Goh PK, Doyle LR, Glass L, Jones KL, Riley EP, Coles CD, Hoyme HE, Kable JA, May PA, Kalberg WO, Sowell ER, Wozniak JR, Mattson SN (2016) A Decision Tree to Identify Children Affected by Prenatal Alcohol Exposure. J Pediatr 177:121–127.e1. [PubMed: 27476634]

Gomez DA, Abdul-Rahman OA (2021) Fetal alcohol spectrum disorders: current state of diagnosis and treatment. Curr Opin Pediatr 33.

Hosmer DW Jr, Lemeshow S, Sturdivant RX (2013) Applied logistic regression John Wiley & Sons.

Hoyme HE, Kalberg WO, Elliott AJ, Blankenship J, Buckley D, Marais A-S, Manning MA, Robinson LK, Adam MP, Abdul-Rahman O, Jewett T, Coles CD, Chambers C, Jones KL, Adnams CM, Shah PE, Riley EP, Charness ME, Warren KR, May PA (2016) Updated Clinical Guidelines for Diagnosing Fetal Alcohol Spectrum Disorders. Pediatrics 138:e20154256–e20154256. [PubMed: 27464676]

Hyland MT, Courchesne-Krak NS, Sobolewski C, Zambrano C, Mattson SN (2023) Neuropsychological Outcomes in FASD Across the Lifespan In: Fetal Alcohol Spectrum Disorders: A Multidisciplinary Approach (Petrenko C, Abdul-Rahman OA eds), pp 221–240. Springer International Publishing.

Jones KL (2011) The effects of alcohol on fetal development. Birth Defects Res Part C - Embryo Today Rev 93:3–11.

Kable JA, Coles CD (2018) Evidence Supporting the Internal Validity of the Proposed ND-PAE Disorder. Child Psychiatry Hum Dev 49:163–175. [PubMed: 28634776]

Kable JA, Coles CD, Jones KL, Yevtushok L, Kulikovsky Y, Zymak-Zakutnya N, Dubchak I, Akhmedzhanova D, Wertelecki W, Chambers CD, CIFASD (2021) Infant Cardiac Orienting Responses Predict Later FASD in the Preschool Period. Alcohol Clin Exp Res 45:386–394. [PubMed: 33277942]

Kable JA, Jones KL (2023) Identifying Prenatal Alcohol Exposure and Children Affected by It: A Review of Biomarkers and Screening Tools. Alcohol Res 43:3.

Kalberg WO, May PA, Buckley D, Hasken JM, Marais AS, De Vries MM, Bezuidenhout H, Manning MA, Robinson LK, Adam MP, Hoyme DB, Parry CDH, Seedat S, Elliott AJ, Hoyme HE (2019) Early-life predictors of fetal alcohol spectrum disorders. Pediatrics 144:e20182141. [PubMed: 31744890]

Kingma DP, Welling M (2019) An Introduction to Variational Autoencoders. Found Trends® Mach Learn 12:307–392.

Kosmicki JA, Sochat V, Duda M, Wall DP (2015) Searching for a minimal set of behaviors for autism detection through feature selection-based machine learning. Transl Psychiatry 5:e514. [PubMed: 25710120]

Kovács G (2019) An empirical comparison and evaluation of minority oversampling techniques on a large number of imbalanced datasets. Appl Soft Comput 83:105662.

LaFrance M-A, McLachlan K, Nash K, Andrew G, Loock C, Oberlander TF, Koren G, Rasmussen C (2014) Evaluation of the neurobehavioral screening tool in children with fetal alcohol spectrum disorders (FASD). J Popul Ther Clin Pharmacol = J la Ther des Popul la Pharmacol Clin 21:e197–210.

Lange S, Probst C, Gmel G, Rehm J, Burd L, Popova S (2017a) Global Prevalence of Fetal Alcohol Spectrum Disorder Among Children and Youth: A Systematic Review and Meta-analysis. JAMA Pediatr 171:948–956. [PubMed: 28828483]

Lange S, Rovet J, Rehm J, Popova S (2017b) Neurodevelopmental profile of Fetal Alcohol Spectrum Disorder: A systematic review. BMC Psychol 5:22. [PubMed: 28645298]

Lever J, Krzywinski M, Altman N (2016) Classification evaluation. Nat Methods 13:603–604.

Little G, Beaulieu C (2020) Multivariate models of brain volume for identification of children and adolescents with fetal alcohol spectrum disorder. Hum Brain Mapp 41:1181–1194. [PubMed: 31737980]

Liu X, Kayser M, Kushner SA, Tiemeier H, Rivadeneira F, Jaddoe VWV, Niessen WJ, Wolvius EB, Roshchupkin GV (2023) Association between prenatal alcohol exposure and children's facial shape: a prospective population-based cohort study. Hum Reprod 38:961–972. [PubMed: 36791805]

Loh HW, Ooi CP, Barua PD, Palmer EE, Molinari F, Acharya UR (2022) Automated detection of ADHD: Current trends and future perspective. Comput Biol Med 146:105525. [PubMed: 35468405]

Lussier AA, Morin AM, MacIsaac JL, Salmon J, Weinberg J, Reynolds JN, Pavlidis P, Chudley AE, Kobor MS (2018) DNA methylation as a predictor of fetal alcohol spectrum disorder. Clin Epigenetics 10:1–14. [PubMed: 29312470]

Lussier AA, Weinberg J, Kobor MS (2017) Epigenetics studies of fetal alcohol spectrum disorder: where are we now? Epigenomics 9:291–311. [PubMed: 28234026]

Mahnke AH, Sideridis GD, Salem NA, Tseng AM, Carter RC, Dodge NC, Rathod AB, Molteno CD, Meintjes EM, Jacobson SW, Miranda RC, Jacobson JL (2021) Infant circulating MicroRNAs as biomarkers of effect in fetal alcohol spectrum disorders. Sci Rep 11:1429. [PubMed: 33446819]

Mattson SN, Jones KL, Chockalingam G, Wozniak JR, Hyland MT, Courchesne-Krak NS, Del Campo M, Riley EP, CIFASD the (2023) Validation of the FASD-Tree as a screening tool for fetal alcohol spectrum disorders. Alcohol Clin Exp Res 47:263–272.

Mattson SN, Roesch SC, Fagerlund A, Autti-Rämö I, Jones KL, May PA, Adnams CM, Konovalova V, Riley EP (2010) Toward a neurobehavioral profile of fetal alcohol spectrum disorders. Alcohol Clin Exp Res 34:1640–1650. [PubMed: 20569243]

May PA, Chambers CD, Kalberg WO, Zellner J, Feldman H, Buckley D, Kopald D, Hasken JM, Xu R, Honerkamp-Smith G, Taras H, Manning MA, Robinson LK, Adam MP, Abdul-Rahman O, Vaux K, Jewett T, Elliott AJ, Kable JA, Akshoomoff N, Falk D, Arroyo JA, Hereld D, Riley EP, Charness ME, Coles CD, Warren KR, Jones KL, Hoyme HE, Daniel F, Arroyo JA, Hereld D, Riley EP, Charness ME, Coles CD, Warren KR, Jones KL, Hoyme HE (2018) Prevalence of fetal alcohol spectrum disorders in 4 US communities. JAMA - J Am Med Assoc 319:474–482.

Mesa DA, Kable JA, Coles CD, Jones KL, Yevtushok L, Kulikovsky Y, Wertelecki W, Coleman TP, Chambers CD, CIFASD (2017) The use of cardiac orienting responses as an early and scalable

biomarker of alcohol-related neurodevelopmental impairment. Alcohol Clin Exp Res 41:128–138. [PubMed: 27883195]

Moons KGM, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG, Woodward M (2012a) Risk prediction models: II. External validation, model updating, and impact assessment. Heart 98:691 LP–698. [PubMed: 22397946]

Moons KGM, Kengne AP, Woodward M, Royston P, Vergouwe Y, Altman DG, Grobbee DE (2012b) Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. Heart 98:683–690. [PubMed: 22397945]

Nash K, Rovet J, Greenbaum R, Fantus E, Nulman I, Koren G (2006) Identifying the behavioural phenotype in fetal alcohol spectrum disorder: sensitivity, specificity and screening potential. Arch Womens Ment Health 9:181–186. [PubMed: 16673042]

Nilsson NJ (2005) Introduction to Machine Learning (No. Copyright 2005 Nils J. Nilsson), Machine Learning

Olson H, Montague R (2011) Prenatal Alcohol Use and FASD: Diagnosis, Assessment and New Directions in Research and Multimodal Treatment, 2011, 64–107 CHAPTER 4 An Innovative Look at Early Intervention for Children Affected by Prenatal Alcohol Exposure. Prenat alcohol use fetal alcohol Spectr Disord diagnosis, Assess new Dir Res multimodal Treat 64–107.

Petrenko CLM, Tahir N, Mahoney EC, Chin NP (2014) Prevention of secondary conditions in fetal alcohol spectrum disorders: identification of systems-level barriers. Matern Child Health J 18:1496–1505. [PubMed: 24178158]

Poldrack RA, Huckins G, Varoquaux G (2020) Establishment of Best Practices for Evidence for Prediction: A Review. JAMA psychiatry 77:534–540. [PubMed: 31774490]

Portales-Casamar E, Lussier AA, Jones MJ, MacIsaac JL, Edgar RD, Mah SM, Barhdadi A, Provost S, Lemieux-Perreault L-P, Cynader MS, Chudley AE, Dubé M-P, Reynolds JN, Pavlidis P, Kobor MS (2016) DNA methylation signature of human fetal alcohol spectrum disorder. Epigenetics Chromatin 9:25. [PubMed: 27358653]

Roomaney I, Nyirenda C, Chetty M (2022) Facial imaging to screen for fetal alcohol spectrum disorder: A scoping review. Alcohol Clin Exp Res 46:1166–1180. [PubMed: 35616438]

Sokolov EN (2002) The Orienting Response in Information Processing L. Erlbaum Associates.

Song D-Y, Topriceanu C-C, Ilie-Ablachim DC, Kinali M, Bisdas S (2021) Machine learning with neuroimaging data to identify autism spectrum disorder: a systematic review and meta-analysis. Neuroradiology 63:2057–2072. [PubMed: 34420058]

Streissguth AP, Bookstein FL, Barr HM, Sampson PD, O'Malley K, Young JK (2004) Risk factors for adverse life outcomes in fetal alcohol syndrome and fetal alcohol effects. J Dev Behav Pediatr 25:228–238. [PubMed: 15308923]

Suttie M, Foroud T, Wetherill L, Jacobson JL, Molteno CD, Meintjes EM, Hoyme HE, Khaole N, Robinson LK, Riley EP, Jacobson SW, Hammond P (2013) Facial Dysmorphism Across the Fetal Alcohol Spectrum. Pediatrics 131:e779–e788. [PubMed: 23439907]

Suttie M, Wetherill L, Jacobson SW, Jacobson JL, Hoyme HE, Sowell ER, Coles C, Wozniak JR, Riley EP, Jones KL, Foroud T, Hammond P (2017) Facial curvature detects and explicates ethnic differences in effects of prenatal alcohol exposure. Alcohol Clin Exp Res 41:1471–1483. [PubMed: 28608920]

Suttie M, Wozniak JR, Parnell SE, Wetherill L, Mattson SN, Sowell ER, Kan E, Riley EP, Jones KL, Coles C, Foroud T, Hammond P (2018) Combined Face-Brain Morphology and Associated Neurocognitive Correlates in Fetal Alcohol Spectrum Disorders. Alcohol Clin Exp Res 42:1769–1782. [PubMed: 29935097]

Valentine M, Bihm DCJ, Wolf L, Hoyme HE, May PA, Buckley D, Kalberg W, Abdul-Rahman OA (2017) Computer-Aided Recognition of Facial Attributes for Fetal Alcohol Spectrum Disorders. Pediatrics 140.

Wall DP, Dally R, Luyster R, Jung JY, DeLuca TF (2012a) Use of artificial intelligence to shorten the behavioral diagnosis of autism. PLoS One 7.

Wall DP, Kosmicki J, Deluca TF, Harstad E, Fusaro VA (2012b) Use of machine learning to shorten observation-based screening and diagnosis of autism. Transl Psychiatry 2:e100–8. [PubMed: 22832900]

Wei Q, Cao H, Shi Y, Xu X, Li T (2023) Machine learning based on eye-tracking data to identify Autism Spectrum Disorder: A systematic review and meta-analysis. J Biomed Inform 137:104254. [PubMed: 36509416]

Yasumura A, Omori M, Fukuda A, Takahashi J, Yasumura Y, Nakagawa E, Koike T, Yamashita Y, Miyajima T, Koeda T, Aihara M, Tachimori H, Inagaki M (2017) Applied Machine Learning Method to Predict Children With ADHD Using Prefrontal Cortex Activity: A Multicenter Study in Japan. J Atten Disord 1087054717740632.

Zhang C, Paolozza A, Tseng P-H, Reynolds JN, Munoz DP, Itti L (2019) Detection of Children/Youth With Fetal Alcohol Spectrum Disorder Through Eye Movement, Psychometric, and Neuroimaging Data. Front Neurol 10:80. [PubMed: 30833926]

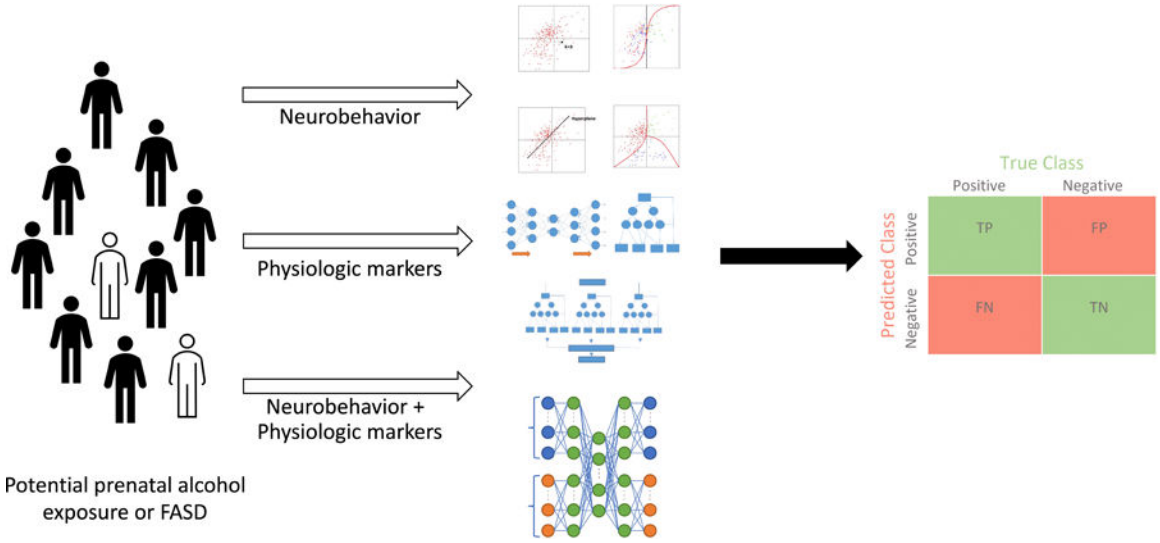
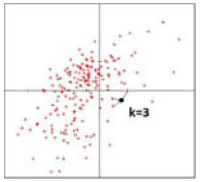
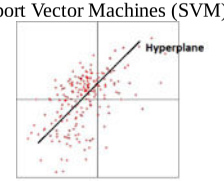
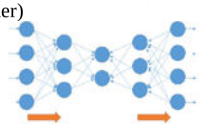
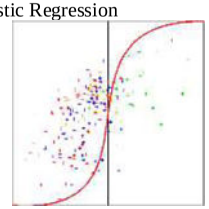


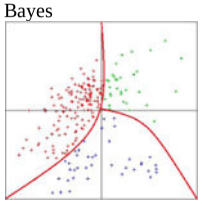
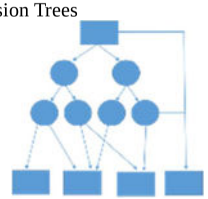
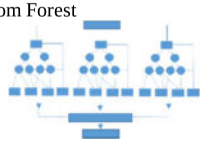
Figure 1:
Schematic of machine learning approaches to screen for prenatal alcohol exposure or FASD

TABLE 1.

Summary of the features of different types of machine learning algorithms.

Algorithm	Type	Description	Advantages/Disadvantages	Uses
K-Nearest 	**Supervised learning**	Classifies data points by comparing them to their k-nearest neighbors, assigning the most common class among them. Reliant on the assumption that data of the same class has similarities.	**+ No training required** **+ Simple implementation** **- Sensitive to scale** **- Doesn't perform well on imbalanced data**	• Binary and Multiclass Classification
Support Vector Machines (SVM) 	**Supervised learning**	Data is transformed into higher dimensional space, where an optimal hyperplane is fitted which best separates the data, while maximizing the margin between them.	**+ Effective for high dimensional data** **- Prone to overfitting** **- Parameter optimization required**	• Binary and Multiclass Classification • Regression (using support vector regression)
Deep Neural Networks (Auto-encoder) 	**Unsupervised learning**	A multi-layered neural network used for unsupervised learning and dimensionality reduction. It consists of an encoder that maps input data to a lower-dimensional latent space and a decoder that attempts to reconstruct the original data from this reduced representation.	**+ Excellent at feature learning** **+ Effective dimensionality reduction** **- Prone to overfitting** **- Computationally expensive and high complexity**	• Dimensionality reduction
Logistic Regression 	**Supervised learning**	Primarily used for binary classification tasks, an S-shaped sigmoid (logistic) function is applied to the data to transform the output of a linear combination of input features into a probability value between 0 and 1.	**+ Simple implementation and computationally efficient** **- Sensitive to outliers** **- Limited to linear decision boundaries**	• Binary Classification

Alcohol Clin Exp Res (Hoboken). Author manuscript; available in PMC 2025 April 01.

Algorithm	Type	Description	Advantages/Disadvantages	Uses
Naïve Bayes 	**Supervised learning**	Uses Bayes' theorem and assumes feature independence. It calculates the probability of a data point belonging to a class based on the joint probabilities of its features, selecting the class with the highest probability.	**+ Simple implementation** **+ Fast** **- Assumes feature independence** **- Struggles with complex data**	• Binary and Multiclass Classification • Regression
Decision Trees 	**Non-parametric supervised learning**	Recursively splits data based on feature attributes to create a tree-like structure. They make decisions by traversing the tree from root to leaf, ultimately assigning data points to a class based on the tree's structure	**+ Fast and computationally efficient** **+ Easily deployed and interpreted** **- Prone to overfitting** **- Sensitive to data variance**	• Binary and Multiclass Classification • Regression
Random Forest 	**Non-parametric supervised learning**	A form of ensemble learning that builds multiple decision trees with bootstrapped data samples and random feature selection. It aggregates their predictions through voting or averaging, reducing overfitting and increasing overall accuracy.	**+ Effective for high dimensional data** **+ Estimates feature importance** **- Computationally expensive and high complexity on large datasets** **- Interpretability issues**	• Binary and Multiclass Classification • Regression