Computer-Aided Recognition of Facial Attributes for Fetal Alcohol Spectrum Disorders

Matthew Valentine, MD,a Dustin C.J. Bihm, MS,a Lior Wolf, PhD,b H. Eugene Hoyme, MD,c,d Philip A. May, PhD,e,f David Buckley, MA, f Wendy Kalberg, MA, LED,f Omar A. Abdul-Rahman, MDg

OBJECTIVES: To compare the detection of facial attributes by computer-based facial recognition software of 2-D images against standard, manual examination in fetal alcohol spectrum disorders (FASD).

METHODS: Participants were gathered from the Fetal Alcohol Syndrome Epidemiology Research database. Standard frontal and oblique photographs of children were obtained during a manual, in-person dysmorphology assessment. Images were submitted for facial analysis conducted by the facial dysmorphology novel analysis technology (an automated system), which assesses ratios of measurements between various facial landmarks to determine the presence of dysmorphic features. Manual blinded dysmorphology assessments were compared with those obtained via the computer-aided system.

RESULTS: Areas under the curve values for individual receiver-operating characteristic curves revealed the computer-aided system (0.88 ± 0.02) to be comparable to the manual method (0.86 ± 0.03) in detecting patients with FASD. Interestingly, cases of alcohol-related neurodevelopmental disorder (ARND) were identified more efficiently by the computer-aided system (0.84 ± 0.07) in comparison to the manual method (0.74 ± 0.04). A facial gestalt analysis of patients with ARND also identified more generalized facial findings compared to the cardinal facial features seen in more severe forms of FASD.

CONCLUSIONS: We found there was an increased diagnostic accuracy for ARND via our computer-aided method. As this category has been historically difficult to diagnose, we believe our experiment demonstrates that facial dysmorphology novel analysis technology can potentially improve ARND diagnosis by introducing a standardized metric for recognizing FASD-associated facial anomalies. Earlier recognition of these patients will lead to earlier intervention with improved patient outcomes.

WHAT'S KNOWN ON THIS SUBJECT: Alcohol-related neurodevelopmental disorder, a category within fetal alcohol spectrum disorders, represents a diagnostic challenge as it lacks apparent morphologic clues to diagnosis. Earlier detection and intervention facilitates improved patient outcomes in fetal alcohol spectrum disorders.

WHAT THIS STUDY ADDS: We report improved efficiency in the detection of patients with alcohol-related neurodevelopmental disorder by using computer-aided methods in comparison with standard, manual methods.

Fetal alcohol spectrum disorders (FASDs) represent an array of structural anomalies and neurocognitive disabilities caused by prenatal alcohol exposure. The harmful nature of prenatal exposure to ethanol is well known, and many mechanisms of teratogenicity have been proposed. Craniofacial development is intimately linked with brain induction and expansion, which explains the association between FASD and certain facial anomalies. Ethanol affects the developing fetus via a diverse range of mechanisms. Importantly, the teratogenic effects of ethanol affect the developing fetus in a temporally dependent manner. Because of this, phenotypic differences in disorders that fall under the FASD umbrella may be evident, even with subtle differences in the stage of prenatal alcohol exposure.2–5

FASD as an entity represents a considerable financial and clinical burden worsened by the fact that many affected individuals go undetected until school age. Late recognition of the etiology for neurocognitive deficits associated with this clinical spectrum leads to poorer developmental outcomes in affected individuals. In addition to primary disabilities, such as decreased cognitive and executive function, adolescents and adults with FASD also face a high rate of secondary disabilities such as problems at school, trouble with the law, alcohol abuse, and mental illness.6 In examining risk factors for adverse life outcomes in FASD, 1 of the strongest correlates of adverse life outcomes, such as disrupted school experiences, is lack of an early diagnosis, with greater delay in diagnosis yielding greater odds of adverse outcomes.7 Consequently, early recognition and diagnosis of FASD should be prioritized by health care professionals aiming to provide adequate support to these patients and minimize the economic impact of the disease.

The range of FASD includes the following diagnostic categories: (1) fetal alcohol syndrome (FAS), (2) partial fetal alcohol syndrome (PFAS), (3) alcohol-related birth defects (ARBDs), and (4) alcohol-related neurodevelopmental disorders (ARNDs).8 Although there are clear diagnostic criteria for FAS, PFAS, and ARBD, a major challenge remains in the effective recognition of patients with ARND because of the relative paucity of specific, easily identifiable features. Although FAS and PFAS can be diagnosed by using physical findings (small head circumference, short palpebral fissures, smooth philtrum, thin vermilion border of the upper lip, and growth deficiency) and neuropsychological evaluation without documented alcohol exposure, ARND diagnostic criteria require documented alcohol exposure because of the lack of physical findings.8 In addition to confirmed maternal alcohol exposure, the diagnostic criteria for ARND that is currently available require recognition of a complex pattern of cognitive and/or behavioral abnormalities inconsistent with developmental level that cannot be explained by genetic predisposition, family background, or environment alone.8–11 As obvious dysmorphic features can be lacking in cases of ARND that do not meet the first criterion, there have been several attempts to accurately define the neurocognitive profile of patients with ARND. The complex nature of defining this neurocognitive profile, however, is demonstrated by the need for advanced cognitive tests involving higher-level processing because IQ tests may be inadequate to differentiate children with ARND from those with developmental disabilities resulting from other causes.9 Despite some success in profiling ARND neurocognitive characteristics, challenges remain in finding a simple and effective tool that can be deployed in the clinical setting to aid in the diagnosis of ARND.

Although the craniofacial clinical criteria for ARND have historically only included reduced head circumference, it has been noted that ARND patients have an increase in minor facial anomalies compared with controls but fewer than those observed in cases of FAS or PFAS.12,13 This suggests there may be some structural findings on physical examination that could be used to aid in the diagnosis of ARND. Evidence exists that computer-based analysis of facial images can detect subclinical features in patients with ARND.14 However, the technology used required complex systems involving three-dimensional cameras that may be impractical in the clinical setting. Therefore, we examined the ability of a commercially available, computer-based facial dysmorphism analysis tool that uses two-dimensional images taken with a standard “point-and-click” camera to aid in the diagnosis of FASD and, specifically, to determine if subclinical facial features in ARND can be detected.

METHODS

Protocols and consent forms were approved by the Human Research Review Committee of the University of New Mexico School of Medicine. Separate active consents for children and mothers to participate were obtained.

Participants in this study were gathered from the Fetal Alcohol Syndrome Epidemiology Research (FASER) database. In FASER studies, children aged 5 to 9 years were evaluated for FASD from locations in South Africa, Italy, and the United States.15–18 The collection of patients in the FASER database relied on active case ascertainment methods to facilitate the efficient identification
of children across the full continuum of FASD diagnoses. The initial sample selection and screening for cases and controls considered deficiencies in growth parameters (≤10th or 25th percentile using height, weight, and head circumference measures) and also random selection of children from the same schools to ensure the representativeness of the sample for the particular population. Subsequently, dysmorphology assessments of all sample children were performed. Standard frontal and oblique photographs were taken of all children. Each child was examined in a blinded manner and independently by 2 dysmorphologists without information about potential prenatal alcohol exposure; affected children were later classified into the appropriate diagnostic category after maternal interviews and neuropsychological testing were conducted in a structured multidisciplinary case conference as previously described. The collected data set of images comprised frontal and oblique photographs of a random subset of patients with FAS from the FASER database: FAS (n = 36), PFAS (n = 31), ARND (n = 22). Sample sizes represent all available cases of FASD (N = 89) present in the FASER database that matched the selection criteria at the time of this study. These were compared with a cohort of control children who were randomly selected and determined to not be affected with FASD and who displayed normal morphology, growth, and development (N = 50).

Appropriate categorical diagnoses were assigned based on specific criteria. A diagnosis of FAS required the following: (1) at least 2 of 3 cardinal facial features (ie, short palpebral fissures, thin vermilion border, smooth philtrum), (2) growth restriction (prenatal and/or postnatal), and (3) evidence of deficient brain growth (presumed by structural brain abnormality and/or occipitofrontal head circumference [OFC] ≤10th percentile). The requirements for a diagnosis of PFAS included the following: (1) at least 2 of 3 cardinal facial features, and (2) prenatal and/or postnatal growth restriction, (3) OFC ≤10th percentile, or (4) a complex behavioral pattern or cognitive abnormalities inconsistent with developmental level. Regarding the cognitive abnormalities, it was required that the observed abnormalities could not be explained by genetic composition (eg, aneuploidies, etc), family history, or environment alone. A preliminary ARND diagnosis required observance of OFC ≤10th percentile or specific behavioral or cognitive abnormalities. Thorough maternal interviews were conducted to confirm maternal alcohol consumption during pregnancy. Confirmed maternal alcohol consumption during pregnancy was required for a diagnosis of ARND. This was in contrast to the diagnosis of cases of FAS and PFAS in which confirmation of maternal alcohol consumption was not strictly required for diagnosis. No cases of ARBD were observed in the cohort.

The Dysmorphology Scoring System

The Dysmorphology Scoring System (DSS) was initially published in 2005 as a research tool by which the dysmorphic features associated with FASD could be quantified in an individual child, thus enabling comparison of overall dysmorphology among affected children. However, the concept was originally envisioned by Dr Jon Aase at the University of New Mexico based on his experience in diagnosing FAS among the general population and Native American populations of New Mexico, Alaska, and other parts of the western United States. The DSS uses a point system from 0 to 3 based on the degree of specificity that a feature has to an FASD diagnosis. For example, each of the 3 cardinal facial features is given a score of 3, whereas “hockey stick” creases and clinodactyly are given a score of 1.

Automated Face Analysis

Automated face analysis was conducted by using the facial dysmorphology novel analysis (FDNA) technology, used in a proprietary software tool called Face2Gene (FDNA Inc, Boston, MA). This technology combines several methodologies of facial recognition from photographs to generate accurate information despite individual differences in variables such as ethnicity, sex, and age. Additionally, this combined methodology maintains superior performance even in the event of skewed image characteristics such as illumination, pose, and expression. The image analysis pathways are demonstrated in Fig 1. First, a subject’s face is detected within the image with a statistical face detector, and the background is discarded. Multiple anatomic points are then located on the face (eg, corners of the eyes and nose, etc). Next, a description of the face under investigation is conducted as multiple lengths, angles, and ratios are computed for each face. These values are then used in combination with statistical mechanisms analyzed statistically to evaluate for the presence of dysmorphic features. Interestingly, because the software analyzes a two-dimensional image, it can make an accurate diagnosis by analyzing the ratios and proportions of the face rather than relying on the features pediatricians most commonly evaluate. The areas that are identified as discriminators include the philtrum and palpebral fissures, although there are other areas that are also recognized as being altered.

In addition to the identification of individual features, we applied a gestalt descriptor to capture and visualize the overall appearance of each face. Because this gestalt
system considers the entire face, it was helpful in reviewing the general appearance of faces of the patients with FASD. The gestalt system encodes the detected face as a compilation of measurements, with each describing the appearance at a specific point of the face. These measurements collectively form a robust template of the face that can be readily compared to other faces. This deep phenotyping technology has also been described as having a role in objectively making diagnoses after a whole-exome analysis.

**Experimental Protocol**

We compared the rate of discrimination achieved by 2 methods: (1) a manual score guided by the DSS with its comprehensive list of manually annotated features and (2) a computer-aided score using the FDNA technology. Four experiments were conducted. In the first 3 experiments, patients with FAS, PFAS, and ARND were considered, and the efficiency of recognition of each specific disorder was evaluated. In the fourth experiment, the 3 groups of disorders were united as a single group before attempting discrimination. In all 4 experiments, the same group of 50 controls was used.

For the second method (FDNA technology), which relies on automatic detections, the list of DSS features used was divided into 2 groups: those detectable in facial images and those that are nonfacial. The nonfacial features were taken from the score sheets completed by experienced pediatric dysmorphologists and used alongside a set of 76 frontal and 39 lateral facial features that were automatically detected. The system then combined the 2 feature types together using a linear classifier whose coefficients were obtained from a separate set of training images put aside for this purpose in a cross-validation manner (see below). The distinction between the 2 groups of features is natural in the clinical setting because most clinicians would not experience difficulty in identifying the nonfacial features; yet challenges in the detection of the facial features can be extensive and often require expertise in dysmorphology.

To estimate the statistical power of the 2 methods to discriminate between patients with FASD and control patients, a cross-validation scheme was used. As with other such schemes, the data were split randomly (multiple times) into training and test sets. Individuals used for training were excluded from the corresponding test set. In our experiments, each such set (train or test) contained half of the examples (both FASD and controls), and the random process was repeated 10 times. At each of the 10 rounds, the parameters of the automatic recognition methods were independently optimized on the training set and evaluated on the test set.

The receiver operating characteristic (ROC) curve is a widely accepted form of measuring classification success. Results for our experiment are reported by computing the area under the curve (AUC) statistic of the ROC curve. An AUC of 1 is indicative of perfect accuracy whereas an AUC of 0.5 is the equivalent performance obtained by a completely random coin-flipping system.

**RESULTS**

The area under the ROC curve for the 4 experiments as well as sensitivity, specificity, and positive and negative predictive values are listed in Table 1. Permutation tests were run on each of the 4 tests (10,000 random tests each) to estimate the probability of obtaining these results in a random manner. To this end, the labels are randomly permuted over the same descriptors 10,000 times, for each of the tests. Thus, the resulting P value shown in Table 2 is the probability of the obtained result over the distribution of random permutations. Individual ROC curves are shown in Fig 2. AUC values for these curves indicated that the semiautomatic method for detecting FASD patients was able to achieve an overall performance comparable to the manual system. Of particular interest, we found cases of ARND to be identified more efficiently by the semiautomatic system than by the manual method.

Visualization of the gestalt analysis is shown in Fig 3. Facial regions colored red represent the primary areas identified in support of an FAS classification, whereas regions less supportive of such a diagnosis are marked with cooler colors. The gestalt analyses were found to be helpful in drawing attention to the most relevant facial regions that support an FASD. In Fig 4, we illustrate gestalt analyses that compare ARND to FAS and PFAS. The average response on patients with ARND is shown for 3 gestalt detectors: ARND, FAS, and PFAS. On visualization, faces of typical patients with ARND appeared to pick up a more diffuse pattern of subtle facial cues in support of classification as ARND. In contrast, FAS and PFAS detectors appeared to focus detection...
### Table 1: AUC Values Obtained by Manual and Computer-Aided Methods for 4 Experiments

<table>
<thead>
<tr>
<th></th>
<th>Manual Score</th>
<th>Computer-Aided Score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AUC ± SD (95% CI)</td>
<td>Sensitivity ± SD (95% CI)</td>
</tr>
<tr>
<td>FAS versus control</td>
<td>0.98 ± 0.03 (0.9414–0.9710)</td>
<td>0.99 ± 0.03 (0.9831–0.9942)</td>
</tr>
<tr>
<td>PFAS versus control</td>
<td>0.89 ± 0.02 (0.8776–0.9024)</td>
<td>0.76 ± 0.16 (0.7571–0.7839)</td>
</tr>
<tr>
<td>ARND versus control</td>
<td>0.74 ± 0.04 (0.7152–0.7648)</td>
<td>0.45 ± 0.24 (0.4402–0.4694)</td>
</tr>
<tr>
<td>Any FASD versus control</td>
<td>0.88 ± 0.02 (0.8676–0.8924)</td>
<td>0.87 ± 0.03 (0.8551–0.8930)</td>
</tr>
</tbody>
</table>

CI, confidence interval.

### Discussion

In searching for improvement in the accuracy of diagnosing FASD, we used the FDNA technology through the commercially available ARND application. Diagnosis of FASD, particularly ARND, remains difficult and there is a clear need for improved efficiency in diagnosing this disorder. We present our experiment as evidence that the computer-aided approach to diagnosing FASD can accurately diagnose ARND and may help clinicians improve the accuracy of diagnosing FASD, particularly ARND. ARND frequently presents our experiment as evidence that facial recognition technology can improve the accuracy of diagnosing FASD, particularly ARND. In addition to the better-recognized dysmorphic facial phenotype, ARND represents a unique phenotype. This spectrum of disorders presents a variety of challenges in diagnosis, including potential for under- or misdiagnoses of FAS, PFAS, ARBD, and ARND. Often under- or misdiagnosed, ARND has historically required the expertise of trained specialists, dysmorphologists in addition to neuropsychological testing conducted by trained experts. Even with the ARND phenotype has not been fully characterized, this spectrum of disorders presents a variety of challenges in diagnosis, including potential for under- or misdiagnoses of FAS, PFAS, ARBD, and ARND. Often under- or misdiagnosed, ARND has historically required the expertise of trained specialists, dysmorphologists in addition to neuropsychological testing conducted by trained experts. Even with the expertise of trained specialists, dysmorphologists in addition to neuropsychological testing conducted by trained experts. Even with the expertise of trained specialists, dysmorphologists, the population of FASD encompasses a variety of disorders, including FAS, PFAS, ARBD, and ARND. FASD encompasses the disease burden among the pediatric population. FASDs represent a significant challenge in diagnosis because, to date, a unique ARND phenotype has not been formulated as opposed to the better-recognized FAS or PFAS phenotype.

### Table 2: Results of Permutation Tests Showing P Values for Each of the 4 Tests

<table>
<thead>
<tr>
<th></th>
<th>PFAS versus control</th>
<th>ARND versus control</th>
<th>Any FASD versus control</th>
<th>FAS versus control</th>
</tr>
</thead>
<tbody>
<tr>
<td>PFAS</td>
<td>0.007</td>
<td>0.007</td>
<td>0.007</td>
<td>0.007</td>
</tr>
<tr>
<td>ARND</td>
<td>0.007</td>
<td>0.007</td>
<td>0.007</td>
<td>0.007</td>
</tr>
<tr>
<td>Any FASD</td>
<td>0.007</td>
<td>0.007</td>
<td>0.007</td>
<td>0.007</td>
</tr>
<tr>
<td>FAS</td>
<td>0.007</td>
<td>0.007</td>
<td>0.007</td>
<td>0.007</td>
</tr>
</tbody>
</table>
available facial recognition and analysis software called Face2Gene and compared efficiency in diagnosis of cases of FASD conducted by this software to a manual method. Our findings revealed that the FDNA technology, when applied to facial assessments, is nearly as accurate as the diagnosis of cases of FAS and PFAS by expert dysmorphologists. Additionally, with our results, we indicate the FDNA technology may actually be more efficient than human dysmorphologists in the assessment and diagnosis of ARND. We describe evidence in support of the idea that subtle facial features exist in patients with ARND who are below the threshold of detection by dysmorphologists. As suggested by Fig 4, the diffuse pattern of structural anomalies detected by FDNA is less discrete for ARBD than for FAS or PFAS; however, general pediatricians, residents, nurses, and other trainees may have difficulty recognizing even the general features of FASD. In cases when prenatal alcohol exposure is documented or suspected, these clinicians can simultaneously improve their ability to categorize dysmorphological features of FASD and use the facial recognition software to screen for children who may need in depth psychological testing for ARND.

This improved detection of patients with ARND by the FDNA system is promising because ARND has been traditionally difficult to diagnose, especially by general pediatric clinicians. Furthermore, the FDNA system relies on two-dimensional image analysis rather than three-dimensional scans. This makes the technology accessible to clinicians because special capturing equipment is unnecessary. This could prove to be particularly important in developing nations where access to molecular testing is limited. Although promising as a detection system, clinical assessment must be used adjunctively with FDNA when characterizing other domains within FASD diagnostic criteria. For example, clinical judgment is needed when assessing the significance of neurodevelopmental delays, the veracity of alcohol exposure reports, and in ruling out other possible genetic or teratogenic causes of the phenotypes identified. We believe further studies examining the diagnosis of ARND using the FDNA technology will provide additional
proof that facial recognition software can improve the accuracy and speed of ARND diagnosis.

If corroborated, our experiments imply that the FDNA technology could play a prominent role in FASD evaluation and provide an accessible tool for clinicians, especially general pediatricians who lack expertise in dysmorphology. The emergence of FDNA as a commercially available software application gives this technology the potential to prove itself as a useful application in the clinician’s toolbox because barriers to accessing this technology remain low for physicians.

CONCLUSIONS

We found increased diagnostic accuracy for ARND via our computer-aided method. Because this category has been historically difficult to diagnose, we believe our experiment demonstrates that technology can improve ARND diagnosis by introducing a standardized metric for recognizing FASD associated facial anomalies. Further improvement in these methods will facilitate earlier recognition of these patients and, ultimately, lead to earlier intervention with improved patient outcomes in a number of areas, including academic, social, and psychological functioning.

ACKNOWLEDGMENTS

We thank the superintendents of schools, administrators, principals, psychologists, and teachers of the school systems in the study communities. They graciously assisted us over the years in the most professional manner. Their support and guidance have been vital to the study.

ABBREVIATIONS

ARBD: alcohol-related birth defect
ARND: alcohol-related neurodevelopmental disorder
AUC: area under the curve
DSS: Dysmorphology Scoring System
FAS: fetal alcohol syndrome
FASD: fetal alcohol spectrum disorder
FASER: Fetal Alcohol Syndrome Epidemiology Research
FDNA: facial dysmorphology novel analysis
OFC: occipitofrontal head circumference
PFAS: partial fetal alcohol syndrome
ROC: receiver operating characteristic

REFERENCES

face-brain dysmorphology patterns are correlative and exposure-stage dependent. *PloS One.* 2012;7(8):e43067


Computer-Aided Recognition of Facial Attributes for Fetal Alcohol Spectrum Disorders
Matthew Valentine, Dustin C.J. Bihm, Lior Wolf, H. Eugene Hoyme, Philip A. May, David Buckley, Wendy Kalberg and Omar A. Abdul-Rahman

*Pediatrics* 2017;140;
DOI: 10.1542/peds.2016-2028 originally published online November 29, 2017;

The online version of this article, along with updated information and services, is located on the World Wide Web at: http://pediatrics.aappublications.org/content/140/6/e20162028