

Original Paper

Training and Profiling a Pediatric Facial Expression Classifier for Children on Mobile Devices: Machine Learning Study

Agnik Banerjee¹; Onur Cezmi Mutlu², MS; Aaron Kline¹, BS; Saimourya Surabhi¹, MS; Peter Washington³, PhD; Dennis Paul Wall^{1,4,5}, PhD

¹Department of Pediatrics (Systems Medicine), Stanford University, Stanford, CA, United States

²Department of Electrical Engineering, Stanford University, Stanford, CA, United States

³Department of Information and Computer Sciences, University of Hawai'i at Mānoa, Honolulu, HI, United States

⁴Department of Biomedical Data Science, Stanford University, Stanford, CA, United States

⁵Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA, United States

Corresponding Author:

Dennis Paul Wall, PhD

Department of Pediatrics (Systems Medicine)

Stanford University

3145 Porter Drive

Stanford, CA, 94304

United States

Phone: 1 650 666 7676

Email: dpwall@stanford.edu

Abstract

Background: Implementing automated facial expression recognition on mobile devices could provide an accessible diagnostic and therapeutic tool for those who struggle to recognize facial expressions, including children with developmental behavioral conditions such as autism. Despite recent advances in facial expression classifiers for children, existing models are too computationally expensive for smartphone use.

Objective: We explored several state-of-the-art facial expression classifiers designed for mobile devices, used posttraining optimization techniques for both classification performance and efficiency on a Motorola Moto G6 phone, evaluated the importance of training our classifiers on children versus adults, and evaluated the models' performance against different ethnic groups.

Methods: We collected images from 12 public data sets and used video frames crowdsourced from the GuessWhat app to train our classifiers. All images were annotated for 7 expressions: neutral, fear, happiness, sadness, surprise, anger, and disgust. We tested 3 copies for each of 5 different convolutional neural network architectures: MobileNetV3-Small 1.0x, MobileNetV2 1.0x, EfficientNetB0, MobileNetV3-Large 1.0x, and NASNetMobile. We trained the first copy on images of children, second copy on images of adults, and third copy on all data sets. We evaluated each model against the entire Child Affective Facial Expression (CAFE) set and by ethnicity. We performed weight pruning, weight clustering, and quantize-aware training when possible and profiled each model's performance on the Moto G6.

Results: Our best model, a MobileNetV3-Large network pretrained on ImageNet, achieved 65.78% accuracy and 65.31% F_1 -score on the CAFE and a 90-millisecond inference latency on a Moto G6 phone when trained on all data. This accuracy is only 1.12% lower than the current state of the art for CAFE, a model with 13.91x more parameters that was unable to run on the Moto G6 due to its size, even when fully optimized. When trained solely on children, this model achieved 60.57% accuracy and 60.29% F_1 -score. When trained only on adults, the model received 53.36% accuracy and 53.10% F_1 -score. Although the MobileNetV3-Large trained on all data sets achieved nearly a 60% F_1 -score across all ethnicities, the data sets for South Asian and African American children achieved lower accuracy (as much as 11.56%) and F_1 -score (as much as 11.25%) than other groups.

Conclusions: With specialized design and optimization techniques, facial expression classifiers can become lightweight enough to run on mobile devices and achieve state-of-the-art performance. There is potentially a "data shift" phenomenon between facial expressions of children compared with adults; our classifiers performed much better when trained on children. Certain underrepresented ethnic groups (e.g., South Asian and African American) also perform significantly worse than groups such as

European Caucasian despite similar data quality. Our models can be integrated into mobile health therapies to help diagnose autism spectrum disorder and provide targeted therapeutic treatment to children.

(*JMIR Form Res* 2023;7:e39917) doi: [10.2196/39917](https://doi.org/10.2196/39917)

KEYWORDS

edge computing; affective computing; autism spectrum disorder; autism; ASD; classifier; classification; model; algorithm; mobile health; computer vision; deep learning; machine learning for health; pediatrics; emotion recognition; mHealth; diagnostic tool; digital therapy; child; developmental disorder; smartphone; image analysis; machine learning; Image classification; neural network

Introduction

Autism spectrum disorder (ASD) affects 1 in 44 children and is the fastest growing developmental disability in the United States [1]. The prevalence of ASD has increased by 61% globally since 2012 [2]. Although research has shown that early detection and therapy are vital for treating ASD [3,4], a lack of access to clinical practitioners, particularly among lower-income families, results in 27% of children over the age of 8 years remaining undiagnosed and too old to respond optimally to treatment [5-8].

Clinicians spend several hours measuring dozens of behavioral features when making a diagnosis [9], further accounting for the long wait times that make it difficult to get an appointment. However, prior research has shown that machine learning models can achieve similar diagnostic capabilities for children with ASD [10-26], providing rapid inference using fewer than 10 behavioral features that can be easily collected through mediums such as short video clips [16,26-29]. Models that analyze a single ASD-related symptom such as speech patterns [30], hand stimming [31], and head banging [32] have provided promising results for diagnosis of ASD when tested on highly heterogeneous data from real children.

Understanding facial expressions is among the most pronounced symptoms for children with ASD, as they often display significant impairments in both the understanding and imitation of facial expression [33,34]. Thus, automated facial expression classifiers can be used to detect ASD by comparing the ability of children to simulate facial expressions compared with neurotypical children in a controlled environment. Additionally, these models can be used for adaptive therapeutic treatment by providing instantaneous feedback to children already diagnosed with ASD who are learning to mimic conventional expressions when exposed to simulated interactions [35-41]. Despite its potential, there have been few endeavors in creating such a model for these purposes, as classifying facial expression is a difficult task. Machine learning models rely on large volumes of data, and children are underrepresented in the few data sets available [42].

To address this issue, we previously developed a mobile game named GuessWhat [43-47], which challenges children with ASD to improve their social interactions while simultaneously collecting structured video data enriched for social human behavior. We subsequently extracted frames from videos recorded by the app during game play and annotated them for the 6 basic emotions described by Ekman and Keltner [48] to create the largest collection of uniquely labeled frames of

children expressing emotion [42]. Using this data set, we trained a facial expression classifier for children that attained state-of-the-art accuracy on the Child Affective Facial Expression (CAFE) data set [49], the standard benchmark in the field for facial expression recognition (FER) of children.

Despite creating this high-performing model, we have yet to leverage it in adaptive digital therapies such as GuessWhat. Due to the decreasing prices of digital technologies and the corresponding widespread availability of mobile devices for almost all socioeconomic levels [50], it is conceivable to use these models in mobile apps, thus offering an alternative medium for autism diagnosis and treatment that is easily accessible and highly affordable. Unfortunately, our prior models were too computationally expensive to successfully run on commercial smartphones, a problem that many other state-of-the-art machine learning models share [51]. However, we hypothesized that it was viable to utilize recent advances in hardware-efficient deep learning architectures to create a facial expression classifier that could be used on mobile devices and be as accurate as these preceding models.

In this study, we evaluated several state-of-the-art expression classifiers designed for use on mobile devices and utilized various posttraining optimization techniques for both classification performance and efficiency on a Motorola Moto G6 phone. We additionally explored the importance of training our classifiers on children rather than adults and evaluated our models against different ethnic groups. Our best model was able to match previous state-of-the-art results on expression recognition for children achieved by Washington et al [42] while being efficient enough to perform inference on the Moto G6 in real time. We highlight the significant performance increase from having children present in the training images and found several ethnic groups that yield worse performance due to being underrepresented. These models can be integrated into mobile health therapies such as the GuessWhat digital health ecosystem to diagnose ASD and provide targeted expression treatment based on the affective profile of the user.

Methods

Ethical Considerations

All study procedures, including data collection, were approved by the Stanford University Institutional Review Board (IRB number 39562) and the Stanford University Privacy Office. In addition, informed consent was obtained from all GuessWhat participants, all of whom had the opportunity to participate in the study without sharing videos.

Data Collection

Because our model was built with the intention of being utilized on mobile devices where videos are often captured from a wide variety of orientations, it was important that the data we used were highly heterogeneous in factors such as lighting and camera angle. Thus, we initially leveraged images from 10 relatively small, yet well-controlled, data sets in order to train our models: National Institute of Mental Health Child Emotional Faces Picture Set (NIMH - ChEFS) [52], Facial Expression Phoenix (FePh) [53], Karolinska Directed Emotional Faces (KDEF) [54], Averaged KDEF (AKDEF) [54], Dartmouth Database of Children's Faces (Dartmouth) [55], Extended Cohn-Kanade Dataset (CK+) [36], Japanese Female Facial Expression (JAFFE) [47], Radboud Faces Dataset (RaFD) [56], NimStim Set of Facial Expressions (NimStim) [57], and the Tsinghua Facial Expression Database (Tsinghua-FED) [58].

Although all these data sets were created in well-controlled environments, they are incredibly diverse when presented in conjunction: NIMH-ChEFS has images from direct and averted gazes; KDEF/AKDEF has images taken from 5 different camera angles; Dartmouth has images taken from 5 different camera

angles and 2 different lighting conditions; CK+ has images taken from frontal and 30 degree views; RaFD has images taken from 3 different gaze directions; and JAFFE, NimStim, and Tsinghua-FED all have images taken from the frontal view.

Because children were severely underrepresented in these data sets and our model is meant to be used with them, we hypothesized that we needed a large data set that focused solely on children's faces. We thus decided to use our data set of images crowdsourced from GuessWhat, which, upon cleaning, contained 21,456 uniquely labeled images of both neurotypical and ASD. We also used a subset of the Face Expression Recognition 2013 (FER-2013) [38] and Expression in-the-Wild (ExpW) [58] data sets, large libraries of web-scraped images, to balance the ratio of samples for each expression. In total, 78,302 images were collected, with approximately 75% of these images consisting of children. This library is roughly as large as the state of the art and follows a similar strategy of using the GuessWhat images in conjunction with external data sets [42]. The participants presented in these data sets also come from a wide array of backgrounds, with detailed demographics (excluding the web-scraped images and FePH, which were not provided) shown in Table 1.

Table 1. Demographics of the training data sets.

Data set ^a	Participants, n	Age (years), mean (SD)	Ethnicity	Female, %	ASD ^b , %
NIMH-ChEFS ^c	59	13.57 (1.66)	East Asian	66.10	0.00
KDEF/AKDEF ^d	70	23.73 (7.24)	Latino	50.00	0.00
Dartmouth ^e	80	9.84 (2.33)	Caucasian	50.00	0.00
CK+ ^f	123	18-50 ^g	81% Caucasian; 13% African American; 6% Other	69.00	0.00
JAFFE ^h	10	N/A ⁱ	East Asian	100.00	0.00
RaFD ^j	49	21.2 (4.0)	Caucasian	51.02	0.00
NimStim ^k	43	19.4 (1.2)	58% Caucasian; 23% Afro-American; 14% East Asian; 5% Latino	41.86	0.00
Tsinghua-FED ^l	Group A: 67; Group B: 70	Group A: 23.82 (4.18); Group B: 64.40 (3.51)	East Asian	Group A: 50.75; Group B: 50.00	0.00
GuessWhat	114	5.98 (2.97)	55.26% Caucasian; 12.28% Hispanic; 9.65% East Asian; 2.63% African; 1.75% Southeast Asian; 1.75% Pacific Islander; 0.87% Arab; 15.81% Unknown	28.95	65.90

^aThe Facial Expression Phoenix (FePH), Face Expression Recognition 2013 (FER-2013), and Expression in-the-Wild (ExpW) data sets were excluded because no demographic details were available.

^bASD: autism spectrum disorder.

^cNIMH-ChEFS: National Institute of Mental Health Child Emotional Faces Picture Set.

^dKDEF/AKDEF: Averaged Karolinska Directed Emotional Faces/Karolinska Directed Emotional Faces.

^eDartmouth Database of Children's Faces.

^fCK+: Extended Cohn-Kanade Dataset.

^gReported as the age range.

^hJAFFE: Japanese Female Facial Expression.

ⁱN/A: not available.

^jRaFD: Radboud Faces Dataset.

^kNimStim Set of Facial Expressions.

^lTsinghua-FED: Tsinghua Facial Expression Database.

As [Table 1](#) shows, Caucasian, Latino, and certain East Asian ethnicities are well represented in nearly all data sets, with a relatively even split among female and male participants. However, African American, Middle Eastern, and South Asian participants are considerably lacking. Furthermore, despite including the GuessWhat images, only ~15% of participants in total had been diagnosed with ASD. Because all children in the evaluation data set are neurotypical, however, this underrepresentation does not pose an issue in this study.

Data Preprocessing

Before training our models, faces were cropped from all images using the Oxford VGGFace model [59] with a ResNet50 backbone. Images were then resized to 224x224 pixels, and grayscale images were converted to 3 color channels. All images were then normalized to a range from -1 to 1.

Model Training

We trained and compared 5 existing architectures designed for use on mobile devices: MobileNetV3-Small 1.0x [60], MobileNetV2 1.0x [61], EfficientNet-B0 [62], MobileNetV3 1.0x [60], and NasNetMobile [63]. All were pretrained on ImageNet [64]. We retrained each layer of each network using categorical cross entropy loss and an Adam optimizer [65] with a learning rate of $1e-5$. During training, all images were subject to a potential horizontal flip, zoomed in or out by a factor up to 0.15, rotated between -45 degrees and 45 degrees, shifted by a factor up to 0.10, and brightened by a factor between 0.80 and 1.20. We assumed the model converged and thus interrupted training once the validation loss did not improve for 5 consecutive epochs.

We trained 3 versions of each model: 1 that included all data sets, 1 that included all data sets that had solely children (NIMH-ChEFS, Dartmouth, and GuessWhat), and 1 that included only adults (KDEF/AKDEF, CK+, JAFFE, RaFD, NimStim, and Tsinghua-FED).

Model Evaluation

We evaluated our models against the CAFE data set [49], a large data set consisting of facial expressions for children, both by ethnicity and in its entirety. CAFE's participants are aged between 2 years and 8 years, the same range in which an ASD diagnosis is most vital ([Table 2](#)) [66]. The child participants in CAFE are from a wide range of racial and ethnic backgrounds, as shown in [Table 3](#). Children in this data set express 7 expressions: happiness, sadness, surprise, fear, anger, disgust, and neutrality.

We evaluated Subset A and Subset B of CAFE to observe our models' performance against faces that human annotators had difficulty classifying [49]. Subset A contains faces that were identified with $\geq 60\%$ accuracy by 100 adult participants. In contrast, Subset B contains faces with substantially greater variability for each expression, resulting in a Cronbach alpha internal consistency score that is 0.052 lower than that of Subset A [49].

We profiled all models on a Motorola Moto G6 Phone using the TensorFlow Lite benchmark application programming interface (API). We also tested our models on an Android demo app we built that performs real-time image classification on a live video feed to ensure our models matched the results from the benchmark tool.

Table 2. Gender distribution by age of the participants in the Child Affective Facial Expression (CAFE) data set.

Age (years)	Gender, n	
	Male	Female
2	0	5
3	11	27
4	126	188
5	210	350
6	86	139
7	0	18
8	7	0

Table 3. Ethnicity of the participants in the Child Affective Facial Expression (CAFE) data set.

Ethnicity	Results, n (%)
Caucasian	519 (43.54)
African American	246 (20.64)
Latino	180 (15.10)
Asian	135 (11.33)
South Asian	112 (9.40)

Model Optimization, Reevaluation, and Profiling

After evaluating on the CAFE data set, we performed weight pruning before fine tuning the network until the validation loss did not improve for 10 consecutive epochs. We then applied weight clustering before fine tuning the network again in an identical fashion. We finally performed quantized-aware training before evaluating the fully optimized model against CAFE. If the model was unable to undergo quantized-aware training, we applied posttraining quantization instead. Once completed, we exported our TensorFlow models to the TensorFlow Lite format and profiled them using the TensorFlow Lite Benchmark framework.

Results

Results on the Entirety of the CAFE Data Set

Upon evaluation, our best model was the MobileNetV3-Large 1.0x that was trained on all data sets, which acquired 65.78%

accuracy and a 65.31% F_1 -score on CAFE (confusion matrix in Figure 1). This performance increased to 78.40% accuracy and a 77.89% F_1 -score on Subset A of CAFE (confusion matrix in Figure 2). When evaluated on CAFE Subset B, the MobileNetV3-Large model acquired 64.77% accuracy and a 65.60% F_1 -score (confusion matrix on Figure 3), attaining accuracies higher than those that even human annotators could achieve [49].

All models except for the NasNetMobile obtained accuracies and F_1 -scores above 61% when trained on all data sets (Table 4), nearly matching state-of-the-art results while being far more efficient [42].

Figure 1. Confusion matrix for the entire Child Affective Facial Expression (CAFE) data set. Each row represents 100%; darker colors represent less frequent occurrences, and lighter colors represent more frequent occurrences, while the true predictions are shown by boxes in the left-to-right diagonal.

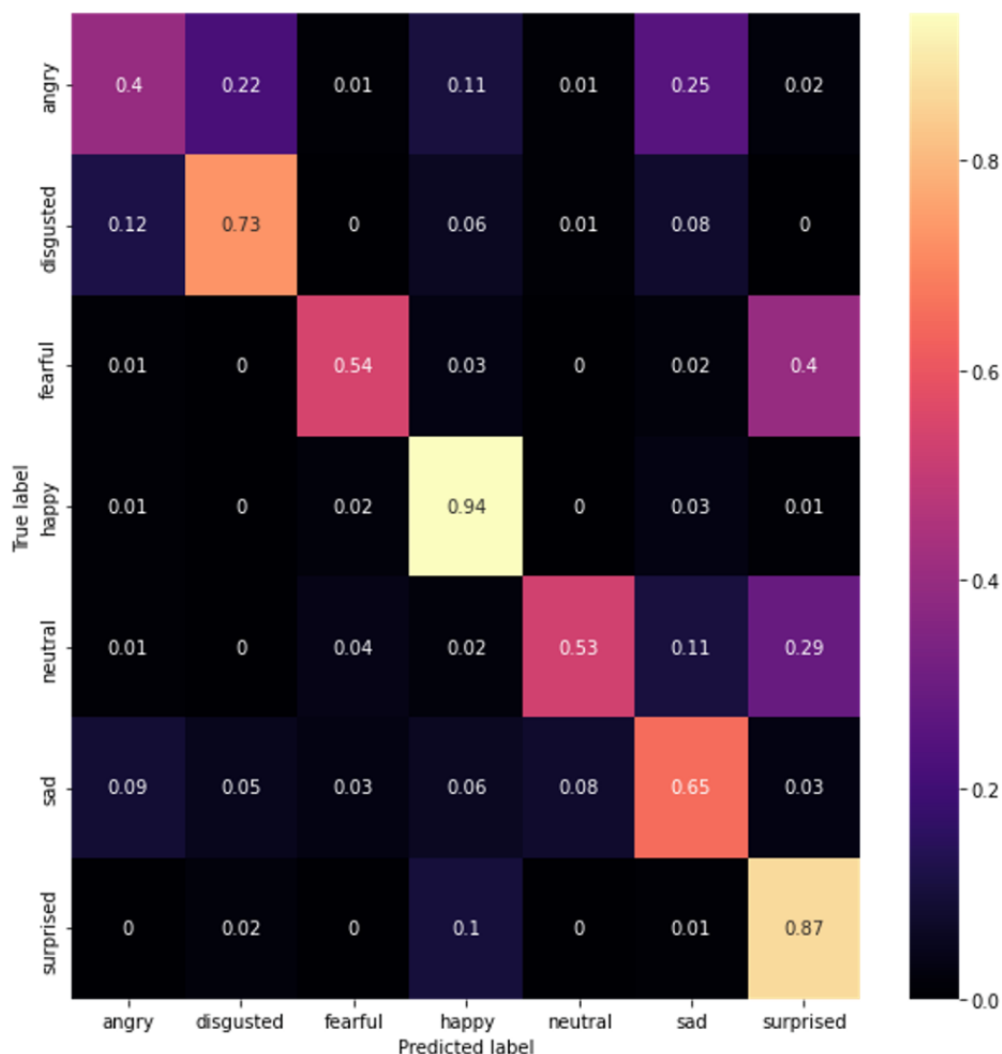


Figure 2. Confusion matrix for Subset A of the Child Affective Facial Expression (CAFE) data set. Each row represents 100%; darker colors represent less frequent occurrences, and lighter colors represent more frequent occurrences, while the true predictions are shown by boxes in the left-to-right diagonal.

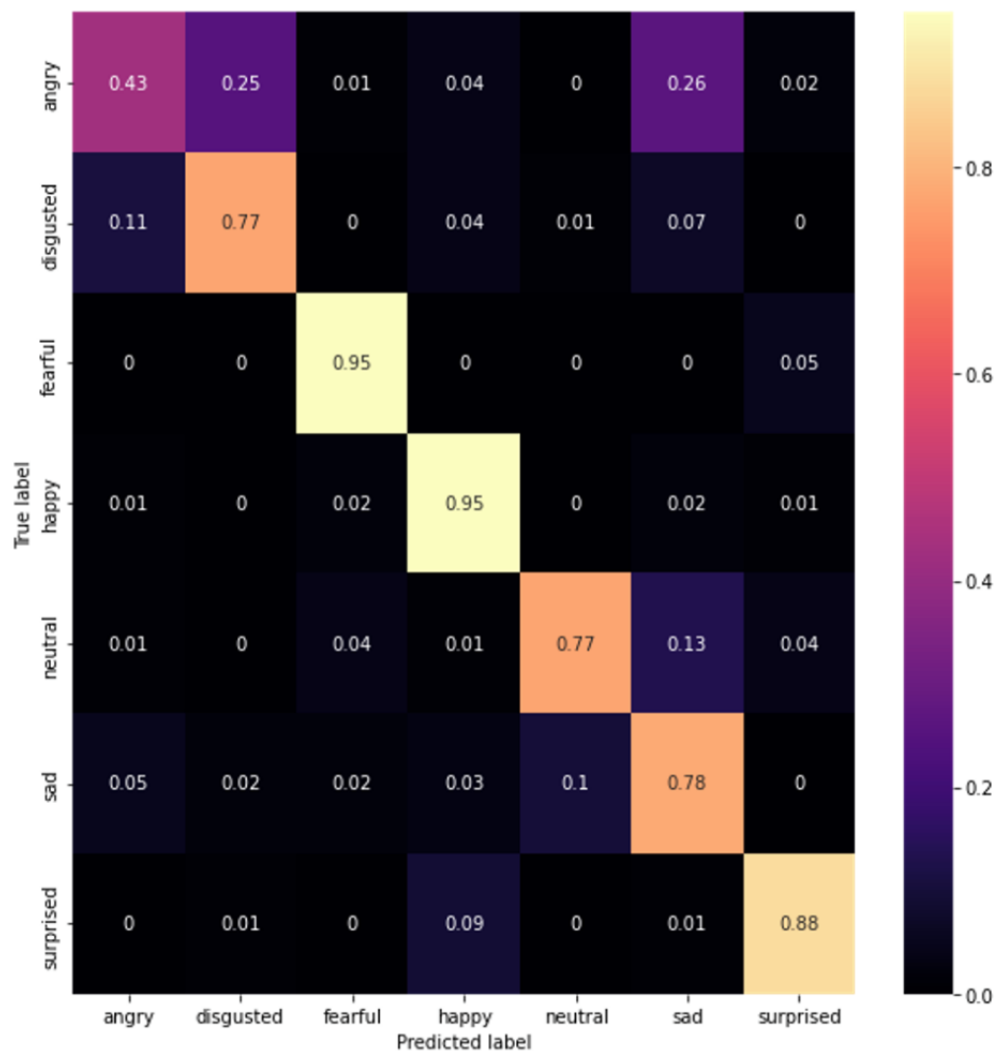


Figure 3. Confusion matrix for Subset B of the Child Affective Facial Expression (CAFE) data set. Each row represents 100%; darker colors represent less frequent occurrences, and lighter colors represent more frequent occurrences, while the true predictions are shown by boxes in the left-to-right diagonal.

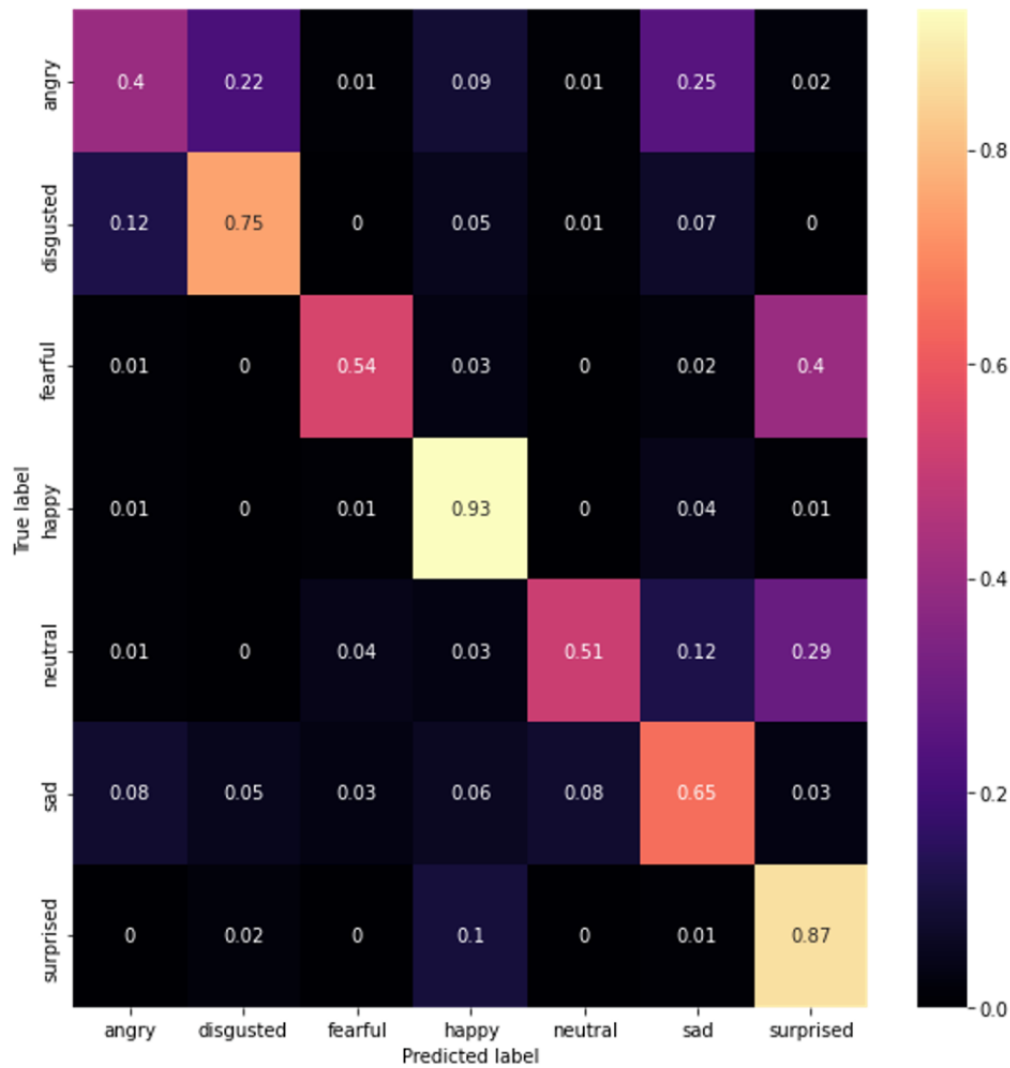


Table 4. Model results when trained on all data sets.

Model	Size Params (in millions)	FLOPs ^a (in millions)	F_1 -score, %		
			Subset A	Subset B	Total
MobileNetV3-Small 1.0x	1.27	56.40	73.68	60.35	61.34
MobileNetV2 1.0x	2.95	300.21	78.04	64.49	63.63
EfficientNetB0	4.74	393.95	75.87	62.75	63.64
MobileNetV3-Large 1.0x	3.52	216.82	76.74	63.55	64.50
NasNetMobile	4.844	568.05	66.82	54.09	55.85

^aFLOPs: floating point operations per second.

CAFE Results When Training on Children Versus Adults Versus All

We evaluated the performance of the best-performing model, the MobileNetV3-Large, when trained on child data sets versus adult data sets. Results are displayed in [Table 5](#).

As shown, training on children yielded better performance than with adults. Although there were nearly 3 times as many images of children that could potentially account for the child model having better performance, the frames in the GuessWhat app are considerably noisy, with a standard of quality clearly worse than that of the adult images that were all collected from well-controlled experiments. Thus, the better performance with poorer data when training on children suggests the importance

of having images of children, even if data are crowdsourced from noisy media such as the GuessWhat app.

Table 5. Comparison of the predictive performance of MobileNetV3-Large.

Group	Accuracy	F_1 -score
Child + adult	0.6578	0.6531
Child	0.6057	0.6029
Adult	0.5336	0.5310

CAFE Results by Ethnicity

Because much of our training data were constrained children of Caucasian, Latino, and East Asian descent, we analyzed the performance of MobileNetV3-Large trained on all data sets against different CAFE subsets. CAFE categorizes its participants into 5 ethnicities: African American, Asian, European American, Latino, and South Asian. Detailed results are shown in Table 6.

As shown in Table 6, the model performed significantly worse on African American and South Asian ethnicities than on other groups, especially European American children, for which our model performed better, achieving as much as 11.56% accuracy and 11.25% F_1 -score better. As shown in Table 1, the very same underrepresented groups had significantly less presence in our training data, indicating that there is a high correlation between the number of training samples and classification performance by ethnicity. Thus, it is reasonable to suggest that, because our training data sets were unbalanced by ethnicity, its performance suffered for underrepresented groups.

Table 6. Results from the MobileNetV3-Large model trained on all data sets when used on different ethnic subsets.

Group	Images, n	Accuracy	F_1 -score
African American	246	0.6127	0.5826
Asian	112	0.6607	0.6648
European American	519	0.6869	0.6745
Latino	180	0.6566	0.6349
South Asian	112	0.5714	0.5620

Performance on an Android Phone

We profiled all 5 models on our Motorola Moto G6 phone and measured the memory consumption and latency when it performed inference on an image. As shown in Table 7, we were able to decrease memory consumption by 4x and latency

by ~1.3x using weight pruning, weight clustering, and quantization without sacrificing accuracy. These improvements are significant considering how few refinements could be made to these specific networks, as they already started out incredibly well-optimized through their well-designed architecture.

Table 7. Latency and memory recorded using 7 CPU threads on the Motorola Moto 6.

Model	Original		Optimized	
	Latency (ms)	Memory (mb)	Latency (ms)	Memory (mb)
MobileNetV3-Small 1.0x	52.33	9.72	45.48	2.77
MobileNetV2 1.0x	62.33	13.78	45.61	4.11
EfficientNetB0	415.04	19.91	301.62	6.08
MobileNetV3-Large 1.0x	124.47	14.82	98.14	4.34
NASNetMobile	218.07	26.98	192.85	8.99

Discussion

Principal Findings

In this study, we trained several machine learning models to recognize expressions on children's faces. We showed the importance of having children in the training data set and that the model performs significantly worse on different ethnicities if they are underrepresented in the training data. Using various optimization techniques, we were able to match state-of-the-art

accuracy while ensuring each model was able to perform real-time inference on a mobile device. We demonstrated that, with specialized training, machine learning models designed to run on edge devices can still match state-of-the-art results on difficult classification tasks.

Limitations

There were a few limitations to this study. Most notably, we only evaluated the performance of our models against CAFE when further evaluation on data sets with more heterogeneity

would better indicate whether the model can generalize on photos taken from mobile devices. Although the GuessWhat data set has these traits, labels from the GuessWhat images are still too noisy after cleaning and need more accurate labels to be a reliable evaluation data set. Another issue is that we were only able to profile our models on a Moto G6 phone. In the future, more extensive testing on devices with less computational power and different operating systems is needed. Although our models performed well on 7 expressions, larger models may be needed to generalize to data with more expressions.

Comparison With Prior Work

The fields of FER and edge machine learning are vast. Prior work relevant to this study can be divided into 3 categories: (1) FER, (2) neural architecture search (NAS), and (3) model compression techniques.

Facial Expression Recognition for ASD Diagnosis and Treatment

FER is a widely researched field with a large library of data sets and classifiers. Early techniques introduced by Kharat and Dudul [35] involved extracting key facial points from faces and passing them through standard models such as support vector machines. Although initial results using this method were promising and computationally inexpensive, these classifiers were evaluated against small, well-structured data sets such as the CK+ [36] and JAFFE [37] data sets. When tested against more heterogeneous data from images taken from a variety of orientations, such as the FER-2013 data set [38], models received much lower scores [39].

Convolutional neural networks (CNNs) have shown the greatest potential in both accuracy and generalizability due to their powerful automatic feature extraction [40,41]. Thus, CNNs are presently the most widely used technique in FER, with an ensemble of CNNs with residual masking blocks leveraged by Pham et al [67] to achieve current state-of-the-art results.

Although results with CNNs in FER have improved consistently through recent years and CNNs have been used in similar applications such as eye gaze detection [68,69], there are few endeavors involving classification on children's faces. The CAFE data set [49] currently is the largest publicly available data set of facial expressions from children and is a standard benchmark in the field of FER on children. The current state of the art on this data set, attained by Washington et al [42], reached 69% accuracy using a ResNet152-V2 architecture pretrained on ImageNet weights and was fine-tuned using data curated from the GuessWhat digital therapy system [43-47].

Prior work has shown that children with ASD are significantly less accurate, need far more time, and require further prompts to respond to facial expression understanding tasks when compared with neurotypical children [70-77]. When mirroring real-life interactions, other studies found that children with ASD especially struggled to understand complex and dynamically displayed expressions, often failing in situations that required fast expression extraction mechanisms [70,73]. The evocation of expressions could also be helpful in detecting whether a child has ASD, with Banire et al [78] discovering that, during controlled experiments, children with ASD often made

expressions such as pressing their lips, something that neurotypical children could not do. Thus, by analyzing the performance of both the understanding and imitation of facial expressions, prior works have shown that facial expression can provide a sensitive biomarker for diagnosing ASD [70-78].

Past studies have also successfully explored using machine learning and other technologies in gamified environments to provide assisted therapy for children with ASD to understand facial expressions with long-term retention [79-83]. Notably, Li et al [80] built a robot-tablet system that offered children the opportunity to play several digital games that practiced their abilities to recognize and imitate facial expressions. Using computer vision and reinforcement learning techniques to predict a child with ASD's facial expression and adjust game strategy to enhance interactive effects, the robot had great therapeutic effect, significantly improving social awareness, cognition, communication, and motivation [80].

Neural Architecture Search

NAS is a paradigm for automating network architecture engineering. NAS can be used to find efficient deep neural network architectures that can be used for FER. Although NAS requires a large amount of computational power to find the optimal network, it can be easily tailored to find the best model for a specific use case. For instance, Lee et al [68] used NAS to build EmotionNet Nano, which was able to outperform other state-of-the-art models at FER while optimizing for speed and energy. Although we did not pursue NAS in this study, we highlight this field as a promising family of methods for mobile model optimization, especially when paired with existing model compression techniques.

Model Compression Techniques

With CNNs being both computationally and memory intensive, several model compression techniques have been developed. Han et al [69] proposed 3 techniques to increase inference speed while decreasing memory overhead and energy consumption: weight pruning, weight clustering, and quantization.

Weight pruning involves gradually zeroing the magnitudes of the weights, making the model sparser by effectively removing weights that have the least significance in the model's predictions. When weight pruning is used with weight clustering, which groups homogeneous weights together to share common values, model size can be decreased by as much as 9 times to 13 times with negligible accuracy loss [69]. By quantizing the standard 32-bit weights of a model to a lower bit representation, models can be further compressed and used in specialized edge hardware for faster inference [84]. In this study, we used all these techniques in conjunction to improve our models' performance.

Conclusions

These models are sufficiently optimized and performant to be used in mobile health therapies such as the GuessWhat smartphone application [43]. GuessWhat delivers therapy by providing important social skill development to children with ASD. Children are exposed to a series of cues and are prompted to respond with the appropriate facial expression—the next

prompt only appears once a caregiver confirms that the child has successfully completed the previous task. By playing the game, children learn to identify which expressions to exhibit under different social contexts while concurrently improving their own execution of these faces. Although these learning exchanges between parent and child are often fruitful, they depend on how well the parents can conduct the session while ensuring that the child is correctly displaying the expressions. This is problematic for parents who do not assess their children with enough rigor or who themselves are unsure of the suitable reaction to a particular setting. Furthermore, if parents are too busy to conduct sessions, children are unable to have the adequate practice necessary for improvement.

Using emotion classifiers can thus remove the human error and bottleneck of requiring another person in the learning session, as they can classify the acted expressions in real time and provide similar instantaneous feedback. These models can be used to compare the child's proficiency with the typical performance of a neurotypical child to indicate how severely a child is unable to recognize and act out expressions, providing an indication of whether a child may have ASD. More holistic analyses are possible when using facial expression classifiers together with models that analyze other phenotypes such as eye gaze and vocal tone. Creating this ecosystem will make autism diagnosis and treatment much more accessible and affordable

to the public, ensuring that children can get the necessary treatment early enough in their lives to have lasting effects.

Although we deployed these models on mobile devices, they are efficient enough to be transferred to other edge devices. SuperpowerGlass is an autism therapeutic delivered on Google Glass, a wearable optical display that responds to touch and voice commands [85-92]. During sessions, children put on the glasses, which capture faces in a child's field of view, and classify faces in real time, providing analysis in the form of emojis that children often find easier to understand. Several game modes are provided to help children learn how to better understand facial expressions. The models we developed can be integrated into this ecosystem to increase the performance of the emotion classifier running on SuperpowerGlass while decreasing power consumption and system performance.

An opportunity for future work includes further recording children's faces from communities that are still heavily underrepresented in facial expression data sets, resulting in lower performance than for other groups in this study. Public data sets with more children are needed. Now that these models can provide inference on mobile devices, another area of promise is integrating these models into on-device training workflows. Once this is complete, federated learning techniques can further improve the models in a privacy-preserving manner while simultaneously providing diagnosis and treatment of ASD.

Acknowledgments

The work was supported in part by funds to DPW from the National Institutes of Health (1R01EB025025-01, 1R01LM013364-01, 1R21HD091500-01, 1R01LM013083); the National Science Foundation (Award 2014232); The Hartwell Foundation; the Bill and Melinda Gates Foundation; the Coulter Foundation; the Lucile Packard Foundation; the Auxiliaries Endowment; the Islamic Development Bank Transform Fund; the Weston Havens Foundation; program grants from Stanford's Human Centered Artificial Intelligence Program, Precision Health and Integrated Diagnostics Center, Beckman Center, Bio-X Center, Predictives and Diagnostics Accelerator, Spectrum, Spark Program in Translational Research, and MediaX; and the Wu Tsai Neurosciences Institute's Neuroscience:Translate Program. We also acknowledge generous support from David Orr, Imma Calvo, Bobby Dekesyer, and Peter Sullivan. PW would like to acknowledge support from Mr. Schroeder and the Stanford Interdisciplinary Graduate Fellowship (SIGF) as the Schroeder Family Goldman Sachs Graduate Fellow.

Conflicts of Interest

DPW is the founder of Cognoa.com. This company is developing digital health solutions for pediatric health care. AK and PW have previously worked as consultants for Cognoa.com. All other authors declare no competing interests.

References

1. Maenner MJ, Shaw KA, Bakian AV, Bilder DA, Durkin MS, Esler A, et al. Prevalence and characteristics of autism spectrum disorder among children aged 8 years - Autism and Developmental Disabilities Monitoring Network, 11 sites, United States, 2018. *MMWR Surveill Summ* 2021 Dec 03;70(11):1-16 [FREE Full text] [doi: [10.15585/mmwr.ss7011a1](https://doi.org/10.15585/mmwr.ss7011a1)] [Medline: [34855725](https://pubmed.ncbi.nlm.nih.gov/34855725/)]
2. Zeidan J, Fombonne E, Scora J, Ibrahim A, Durkin MS, Saxena S, et al. Global prevalence of autism: A systematic review update. *Autism Res* 2022 May;15(5):778-790 [FREE Full text] [doi: [10.1002/aur.2696](https://doi.org/10.1002/aur.2696)] [Medline: [35238171](https://pubmed.ncbi.nlm.nih.gov/35238171/)]
3. Dawson G. Early behavioral intervention, brain plasticity, and the prevention of autism spectrum disorder. *Dev Psychopathol* 2008 Jul 07;20(3):775-803. [doi: [10.1017/s0954579408000370](https://doi.org/10.1017/s0954579408000370)]
4. Landa RJ, Holman KC, Garrett-Mayer E. Social and communication development in toddlers with early and later diagnosis of autism spectrum disorders. *Arch Gen Psychiatry* 2007 Jul;64(7):853-864. [doi: [10.1001/archpsyc.64.7.853](https://doi.org/10.1001/archpsyc.64.7.853)] [Medline: [17606819](https://pubmed.ncbi.nlm.nih.gov/17606819/)]
5. Battle D. Diagnostic and Statistical Manual of Mental Disorders (DSM). *Codas* 2013;25(2):191-192 [FREE Full text] [doi: [10.1590/s2317-17822013000200017](https://doi.org/10.1590/s2317-17822013000200017)] [Medline: [24413388](https://pubmed.ncbi.nlm.nih.gov/24413388/)]

6. Kogan M, Vladutiu C, Schieve L, Ghandour R, Blumberg S, Zablotsky B, et al. The prevalence of parent-reported autism spectrum disorder among US children. *Pediatrics* 2018 Dec;142(6):1 [FREE Full text] [doi: [10.1542/peds.2017-4161](https://doi.org/10.1542/peds.2017-4161)] [Medline: [30478241](https://pubmed.ncbi.nlm.nih.gov/30478241/)]
7. Mazurek M, Handen B, Wodka E, Nowinski L, Butter E, Engelhardt C. Age at first autism spectrum disorder diagnosis: the role of birth cohort, demographic factors, and clinical features. *J Dev Behav Pediatr* 2014;35(9):561-569. [doi: [10.1097/DBP.0000000000000097](https://doi.org/10.1097/DBP.0000000000000097)] [Medline: [25211371](https://pubmed.ncbi.nlm.nih.gov/25211371/)]
8. Siklos S, Kerns KA. Assessing the diagnostic experiences of a small sample of parents of children with autism spectrum disorders. *Res Dev Disabil* 2007 Jan;28(1):9-22. [doi: [10.1016/j.ridd.2005.09.003](https://doi.org/10.1016/j.ridd.2005.09.003)] [Medline: [16442261](https://pubmed.ncbi.nlm.nih.gov/16442261/)]
9. Lord C, Rutter M, Le Couteur A. Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J Autism Dev Disord* 1994 Oct;24(5):659-685. [doi: [10.1007/BF02172145](https://doi.org/10.1007/BF02172145)] [Medline: [7814313](https://pubmed.ncbi.nlm.nih.gov/7814313/)]
10. Leblanc E, Washington P, Varma M, Dunlap K, Penev Y, Kline A, et al. Feature replacement methods enable reliable home video analysis for machine learning detection of autism. *Sci Rep* 2020 Dec 04;10(1):21245 [FREE Full text] [doi: [10.1038/s41598-020-76874-w](https://doi.org/10.1038/s41598-020-76874-w)] [Medline: [33277527](https://pubmed.ncbi.nlm.nih.gov/33277527/)]
11. Washington P, Park N, Srivastava P, Voss C, Kline A, Varma M, et al. Data-driven diagnostics and the potential of mobile artificial intelligence for digital therapeutic phenotyping in computational psychiatry. *Biol Psychiatry Cogn Neurosci Neuroimaging* 2020 Aug;5(8):759-769 [FREE Full text] [doi: [10.1016/j.bpsc.2019.11.015](https://doi.org/10.1016/j.bpsc.2019.11.015)] [Medline: [32085921](https://pubmed.ncbi.nlm.nih.gov/32085921/)]
12. Parikh MN, Li H, He L. Enhancing diagnosis of autism with optimized machine learning models and personal characteristic data. *Front Comput Neurosci* 2019 Feb 15;13:9 [FREE Full text] [doi: [10.3389/fncom.2019.00009](https://doi.org/10.3389/fncom.2019.00009)] [Medline: [30828295](https://pubmed.ncbi.nlm.nih.gov/30828295/)]
13. Washington P, Leblanc E, Dunlap K, Penev Y, Kline A, Paskov K, et al. Precision telemedicine through crowdsourced machine learning: testing variability of crowd workers for video-based autism feature recognition. *J Pers Med* 2020 Aug 13;10(3):86 [FREE Full text] [doi: [10.3390/jpm10030086](https://doi.org/10.3390/jpm10030086)] [Medline: [32823538](https://pubmed.ncbi.nlm.nih.gov/32823538/)]
14. Washington P, Paskov K, Kalantarian H, Stockham N, Voss C, Kline A, et al. Feature selection and dimension reduction of social autism data. *Pac Symp Biocomput* 2020;25:707-718 [FREE Full text] [Medline: [31797640](https://pubmed.ncbi.nlm.nih.gov/31797640/)]
15. Washington P, Kalantarian H, Tariq Q, Schwartz J, Dunlap K, Chrisman B, et al. Validity of online screening for autism: crowdsourcing study comparing paid and unpaid diagnostic tasks. *J Med Internet Res* 2019 May 23;21(5):e13668. [doi: [10.2196/13668](https://doi.org/10.2196/13668)]
16. Tariq Q, Fleming SL, Schwartz JN, Dunlap K, Corbin C, Washington P, et al. Detecting developmental delay and autism through machine learning models using home videos of Bangladeshi children: development and validation study. *J Med Internet Res* 2019 Apr 24;21(4):e13822 [FREE Full text] [doi: [10.2196/13822](https://doi.org/10.2196/13822)] [Medline: [31017583](https://pubmed.ncbi.nlm.nih.gov/31017583/)]
17. Washington P, Leblanc E, Dunlap K, Penev Y, Varma M, Jung JY, et al. Selection of trustworthy crowd workers for telemedical diagnosis of pediatric autism spectrum disorder. *Pac Symp Biocomput* 2021;26:14-25 [FREE Full text] [Medline: [33691000](https://pubmed.ncbi.nlm.nih.gov/33691000/)]
18. Washington P, Tariq Q, Leblanc E, Chrisman B, Dunlap K, Kline A, et al. Crowdsourced privacy-preserved feature tagging of short home videos for machine learning ASD detection. *Sci Rep* 2021 Apr 07;11(1):7620 [FREE Full text] [doi: [10.1038/s41598-021-87059-4](https://doi.org/10.1038/s41598-021-87059-4)] [Medline: [33828118](https://pubmed.ncbi.nlm.nih.gov/33828118/)]
19. Washington P, Chrisman B, Leblanc E, Dunlap K, Kline A, Mutlu C, et al. Crowd annotations can approximate clinical autism impressions from short home videos with privacy protections. *Intell Based Med* 2022;6:100056 [FREE Full text] [doi: [10.1016/j.ibmed.2022.100056](https://doi.org/10.1016/j.ibmed.2022.100056)] [Medline: [35634270](https://pubmed.ncbi.nlm.nih.gov/35634270/)]
20. Varma M, Washington P, Chrisman B, Kline A, Leblanc E, Paskov K, et al. Identification of social engagement indicators associated with autism spectrum disorder using a game-based mobile app: comparative study of gaze fixation and visual scanning methods. *J Med Internet Res* 2022 Feb 15;24(2):e31830 [FREE Full text] [doi: [10.2196/31830](https://doi.org/10.2196/31830)] [Medline: [35166683](https://pubmed.ncbi.nlm.nih.gov/35166683/)]
21. Abbas H, Garberson F, Liu-Mayo S, Glover E, Wall DP. Multi-modular AI approach to streamline autism diagnosis in young children. *Sci Rep* 2020 Mar 19;10(1):5014 [FREE Full text] [doi: [10.1038/s41598-020-61213-w](https://doi.org/10.1038/s41598-020-61213-w)] [Medline: [32193406](https://pubmed.ncbi.nlm.nih.gov/32193406/)]
22. Wall DP, Kosmicki J, Deluca TF, Harstad E, Fusaro VA. Use of machine learning to shorten observation-based screening and diagnosis of autism. *Transl Psychiatry* 2012 Apr 10;2(4):e100-e100 [FREE Full text] [doi: [10.1038/tp.2012.10](https://doi.org/10.1038/tp.2012.10)] [Medline: [22832900](https://pubmed.ncbi.nlm.nih.gov/22832900/)]
23. Duda M, Daniels J, Wall DP. Clinical Evaluation of a Novel and Mobile Autism Risk Assessment. *J Autism Dev Disord* 2016 Jun 12;46(6):1953-1961 [FREE Full text] [doi: [10.1007/s10803-016-2718-4](https://doi.org/10.1007/s10803-016-2718-4)] [Medline: [26873142](https://pubmed.ncbi.nlm.nih.gov/26873142/)]
24. Fusaro VA, Daniels J, Duda M, DeLuca TF, D'Angelo O, Tamburello J, et al. The potential of accelerating early detection of autism through content analysis of YouTube videos. *PLoS One* 2014 Apr 16;9(4):e93533 [FREE Full text] [doi: [10.1371/journal.pone.0093533](https://doi.org/10.1371/journal.pone.0093533)] [Medline: [24740236](https://pubmed.ncbi.nlm.nih.gov/24740236/)]
25. Washington P, Tariq Q, Leblanc E, Chrisman B, Dunlap K, Kline A, et al. Crowdsourced feature tagging for scalable and privacy-preserved autism diagnosis. medRxiv. Preprint posted online December 17, 2020 2020 [FREE Full text] [doi: [10.1101/2020.12.15.20248283](https://doi.org/10.1101/2020.12.15.20248283)]
26. Wall DP, Dally R, Luyster R, Jung J, Deluca TF. Use of artificial intelligence to shorten the behavioral diagnosis of autism. *PLoS One* 2012 Aug 27;7(8):e43855 [FREE Full text] [doi: [10.1371/journal.pone.0043855](https://doi.org/10.1371/journal.pone.0043855)] [Medline: [22952789](https://pubmed.ncbi.nlm.nih.gov/22952789/)]

27. Tariq Q, Daniels J, Schwartz JN, Washington P, Kalantarian H, Wall DP. Mobile detection of autism through machine learning on home video: A development and prospective validation study. *PLoS Med* 2018 Nov 27;15(11):e1002705 [FREE Full text] [doi: [10.1371/journal.pmed.1002705](https://doi.org/10.1371/journal.pmed.1002705)] [Medline: [30481180](https://pubmed.ncbi.nlm.nih.gov/30481180/)]
28. Levy S, Duda M, Haber N, Wall DP. Sparsifying machine learning models identify stable subsets of predictive features for behavioral detection of autism. *Mol Autism* 2017 Dec 19;8(1):65 [FREE Full text] [doi: [10.1186/s13229-017-0180-6](https://doi.org/10.1186/s13229-017-0180-6)] [Medline: [29270283](https://pubmed.ncbi.nlm.nih.gov/29270283/)]
29. Kosmicki JA, Sochat V, Duda M, Wall DP. Searching for a minimal set of behaviors for autism detection through feature selection-based machine learning. *Transl Psychiatry* 2015 Feb 24;5(2):e514-e514 [FREE Full text] [doi: [10.1038/tp.2015.7](https://doi.org/10.1038/tp.2015.7)] [Medline: [25710120](https://pubmed.ncbi.nlm.nih.gov/25710120/)]
30. Chi N, Washington P, Kline A, Husic A, Hou C, He C, et al. Classifying autism from crowdsourced semistructured speech recordings: machine learning model comparison study. *JMIR Pediatr Parent* 2022 Apr 14;5(2):e35406 [FREE Full text] [doi: [10.2196/35406](https://doi.org/10.2196/35406)] [Medline: [35436234](https://pubmed.ncbi.nlm.nih.gov/35436234/)]
31. Lakkapragada A, Kline A, Mutlu O, Paskov K, Chrisman B, Stockham N, et al. The classification of abnormal hand movement to aid in autism detection: machine learning study. *JMIR Biomed Eng* 2022 Jun 6;7(1):e33771. [doi: [10.2196/33771](https://doi.org/10.2196/33771)]
32. Washington P, Kline A, Mutlu OC, Leblanc E, Hou C, Stockham N, et al. Activity Recognition with Moving Cameras and Few Training Examples: Applications for Detection of Autism-Related Headbanging. 2021 Presented at: CHI EA '21: Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems; May 8-13, 2021; Yokohama, Japan. [doi: [10.1145/3411763.3451701](https://doi.org/10.1145/3411763.3451701)]
33. Trevisan DA, Bowering M, Birmingham E. Alexithymia, but not autism spectrum disorder, may be related to the production of emotional facial expressions. *Mol Autism* 2016 Nov 11;7(1):46 [FREE Full text] [doi: [10.1186/s13229-016-0108-6](https://doi.org/10.1186/s13229-016-0108-6)] [Medline: [27895883](https://pubmed.ncbi.nlm.nih.gov/27895883/)]
34. Manfredonia J, Bangerter A, Manyakov NV, Ness S, Lewin D, Skalkin A, et al. Automatic recognition of posed facial expression of emotion in individuals with autism spectrum disorder. *J Autism Dev Disord* 2019 Jan;49(1):279-293. [doi: [10.1007/s10803-018-3757-9](https://doi.org/10.1007/s10803-018-3757-9)] [Medline: [30298462](https://pubmed.ncbi.nlm.nih.gov/30298462/)]
35. Kharat GU, Dudul SV. Human emotion recognition system using optimally designed SVM with different facial feature extraction techniques. *WSEAS Transactions on Computers* 2008;7(6):650-659 [FREE Full text]
36. Lucey P, Cohn JF, Kanade T, Saragih J, Ambadar Z, Matthews I. The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression. 2010 Presented at: IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops; June 13-18, 2020; San Francisco, CA. [doi: [10.1109/cvprw.2010.5543262](https://doi.org/10.1109/cvprw.2010.5543262)]
37. Lyons M, Akamatsu S, Kamachi M, Gyoba J. Coding facial expressions with Gabor wavelets. 2002 Presented at: Third IEEE International Conference on Automatic Face and Gesture Recognition; April 14-16, 1998; Nara, Japan. [doi: [10.1109/afgr.1998.670949](https://doi.org/10.1109/afgr.1998.670949)]
38. Goodfellow I, Erhan D, Luc Carrier P, Courville A, Mirza M, Hamner B, et al. Challenges in representation learning: a report on three machine learning contests. *Neural Netw* 2015 Apr;64:59-63. [doi: [10.1016/j.neunet.2014.09.005](https://doi.org/10.1016/j.neunet.2014.09.005)] [Medline: [25613956](https://pubmed.ncbi.nlm.nih.gov/25613956/)]
39. Pramerdorfer C, Kappel M. Facial expression recognition using convolutional neural networks: state of the art. arXiv. Preprint posted online December 9, 2016 2020.
40. Revina I, Emmanuel WS. A Survey on Human Face Expression Recognition Techniques. *Journal of King Saud University - Computer and Information Sciences* 2021 Jul;33(6):619-628. [doi: [10.1016/j.jksuci.2018.09.002](https://doi.org/10.1016/j.jksuci.2018.09.002)]
41. Ko B. A Brief Review of Facial Emotion Recognition Based on Visual Information. *Sensors (Basel)* 2018 Jan 30;18(2):401 [FREE Full text] [doi: [10.3390/s18020401](https://doi.org/10.3390/s18020401)] [Medline: [29385749](https://pubmed.ncbi.nlm.nih.gov/29385749/)]
42. Washington P, Kalantarian H, Kent J, Husic A, Kline A, Leblanc E, et al. Training an emotion detection classifier using frames from a mobile therapeutic game for children with developmental disorders. arXiv. Preprint posted online December 16, 2020 2020.
43. Kalantarian H, Washington P, Schwartz J, Daniels J, Haber N, Wall DP. Guess What?: Towards understanding autism from structured video using facial affect. *J Healthc Inform Res* 2019 Oct 2;3(1):43-66 [FREE Full text] [doi: [10.1007/s41666-018-0034-9](https://doi.org/10.1007/s41666-018-0034-9)] [Medline: [33313475](https://pubmed.ncbi.nlm.nih.gov/33313475/)]
44. Kalantarian H, Jedoui K, Dunlap K, Schwartz J, Washington P, Husic A, et al. The performance of emotion classifiers for children with parent-reported autism: quantitative feasibility study. *JMIR Ment Health* 2020 Apr 01;7(4):e13174 [FREE Full text] [doi: [10.2196/13174](https://doi.org/10.2196/13174)] [Medline: [32234701](https://pubmed.ncbi.nlm.nih.gov/32234701/)]
45. Kalantarian H, Washington P, Schwartz J, Daniels J, Haber N, Wall D. A gamified mobile system for crowdsourcing video for autism research. 2018 Presented at: IEEE International Conference on Healthcare Informatics (ICHI); June 4-7, 2018; New York, NY. [doi: [10.1109/ichi.2018.00052](https://doi.org/10.1109/ichi.2018.00052)]
46. Kalantarian H, Jedoui K, Washington P, Tariq Q, Dunlap K, Schwartz J, et al. Labeling images with facial emotion and the potential for pediatric healthcare. *Artif Intell Med* 2019 Jul;98:77-86 [FREE Full text] [doi: [10.1016/j.artmed.2019.06.004](https://doi.org/10.1016/j.artmed.2019.06.004)] [Medline: [31521254](https://pubmed.ncbi.nlm.nih.gov/31521254/)]
47. Kalantarian H, Jedoui K, Washington P, Wall DP. A mobile game for automatic emotion-labeling of images. *IEEE Trans Games* 2020 Jun;12(2):213-218 [FREE Full text] [doi: [10.1109/tg.2018.2877325](https://doi.org/10.1109/tg.2018.2877325)] [Medline: [32551410](https://pubmed.ncbi.nlm.nih.gov/32551410/)]

48. Kahn RC, Jascott D, Carlon GC, Schweizer O, Howland WS, Goldiner PL. Massive blood replacement: correlation of ionized calcium, citrate, and hydrogen ion concentration. *Anesth Analg* 1979;58(4):274-278. [doi: [10.1213/00000539-197907000-00003](https://doi.org/10.1213/00000539-197907000-00003)] [Medline: [36816](https://pubmed.ncbi.nlm.nih.gov/36816/)]
49. LoBue V, Thrasher C. The Child Affective Facial Expression (CAFE) set: validity and reliability from untrained adults. *Front Psychol* 2014 Jan 06;5:1532 [FREE Full text] [doi: [10.3389/fpsyg.2014.01532](https://doi.org/10.3389/fpsyg.2014.01532)] [Medline: [25610415](https://pubmed.ncbi.nlm.nih.gov/25610415/)]
50. Frias-Martinez V, Virseda J. On the relationship between socio-economic factors and cell phone usage. 2012 Presented at: ICTD '12: Fifth International Conference on Information and Communication Technologies and Development; March 12-15, 2012; Atlanta, GA. [doi: [10.1145/2160673.2160684](https://doi.org/10.1145/2160673.2160684)]
51. Ignatov A, Timofte R, Chou W, Wang K, Wu M, Hartley T, et al. AI Benchmark: Running Deep Neural Networks on Android Smartphones. In: Leal-Taixé L, Roth S, editors. *Computer Vision – ECCV 2018 Workshops*. ECCV 2018. Lecture Notes in Computer Science, vol 11133. Cham, Switzerland: Springer; 2019.
52. Egger HL, Pine DS, Nelson E, Leibenluft E, Ernst M, Towbin KE, et al. The NIMH Child Emotional Faces Picture Set (NIMH-ChEFS): a new set of children's facial emotion stimuli. *Int J Methods Psychiatr Res* 2011 Sep 24;20(3):145-156 [FREE Full text] [doi: [10.1002/mpr.343](https://doi.org/10.1002/mpr.343)] [Medline: [22547297](https://pubmed.ncbi.nlm.nih.gov/22547297/)]
53. Alagbhand M, Yousefi N, Garibay I. Facial expression phoenix (FePh): an annotated sequenced dataset for facial and emotion-specified expressions in sign language. arXiv. Preprint posted online March 3, 2020 2020.
54. Goeleven E, De Raedt R, Leyman L, Verschuere B. The Karolinska Directed Emotional Faces: A validation study. *Cognition & Emotion* 2008 Sep;22(6):1094-1118. [doi: [10.1080/02699930701626582](https://doi.org/10.1080/02699930701626582)]
55. Dalrymple KA, Gomez J, Duchaine B. The Dartmouth Database of Children's Faces: acquisition and validation of a new face stimulus set. *PLoS One* 2013 Nov 14;8(11):e79131 [FREE Full text] [doi: [10.1371/journal.pone.0079131](https://doi.org/10.1371/journal.pone.0079131)] [Medline: [24244434](https://pubmed.ncbi.nlm.nih.gov/24244434/)]
56. Langner O, Dotsch R, Bijlstra G, Wigboldus DHJ, Hawk ST, van Knippenberg A. Presentation and validation of the Radboud Faces Database. *Cognition & Emotion* 2010 Dec;24(8):1377-1388. [doi: [10.1080/02699930903485076](https://doi.org/10.1080/02699930903485076)]
57. Tottenham N, Tanaka JW, Leon AC, McCarry T, Nurse M, Hare TA, et al. The NimStim set of facial expressions: judgments from untrained research participants. *Psychiatry Res* 2009 Aug 15;168(3):242-249 [FREE Full text] [doi: [10.1016/j.psychres.2008.05.006](https://doi.org/10.1016/j.psychres.2008.05.006)] [Medline: [19564050](https://pubmed.ncbi.nlm.nih.gov/19564050/)]
58. Yang T, Yang Z, Xu G, Gao D, Zhang Z, Wang H, et al. Tsinghua facial expression database - A database of facial expressions in Chinese young and older women and men: Development and validation. *PLoS One* 2020 Apr 15;15(4):e0231304 [FREE Full text] [doi: [10.1371/journal.pone.0231304](https://doi.org/10.1371/journal.pone.0231304)] [Medline: [32294105](https://pubmed.ncbi.nlm.nih.gov/32294105/)]
59. Parkhi O, Vedaldi A, Zisserman A. Deep face recognition. 2015 Presented at: British Machine Vision Conference (BMVC); September 7-10, 2015; Swansea, United Kingdom URL: <http://www.bmva.org/bmvc/2015/papers/paper041/index.html> [doi: [10.5244/c.29.41](https://doi.org/10.5244/c.29.41)]
60. Howard A, Sandler M, Chen B, Wang W, Chen LC, Tan M, et al. Searching for MobileNetV3. 2019 Presented at: IEEE/CVF International Conference on Computer Vision (ICCV); October 27 - November 2, 2019; Seoul, Republic of Korea. [doi: [10.1109/iccv.2019.00140](https://doi.org/10.1109/iccv.2019.00140)]
61. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC. MobileNetV2: Inverted Residuals and Linear Bottlenecks. 2018 Presented at: IEEE/CVF Conference on Computer Vision and Pattern Recognition; June 18-23, 2018; Salt Lake City, UT. [doi: [10.1109/cvpr.2018.00474](https://doi.org/10.1109/cvpr.2018.00474)]
62. Tan M, Le Q. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. 2019 Presented at: International Conference on Machine Learning (ICML); June 9-15, 2019; Long Beach, CA.
63. Zoph B, Vasudevan V, Shlens J, Le QV. Learning Transferable Architectures for Scalable Image Recognition. 2018 Presented at: IEEE/CVF Conference on Computer Vision and Pattern Recognition; June 18-23, 2018; Salt Lake City, UT. [doi: [10.1109/cvpr.2018.00907](https://doi.org/10.1109/cvpr.2018.00907)]
64. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. ImageNet: A large-scale hierarchical image database. 2009 Presented at: IEEE Conference on Computer Vision and Pattern Recognition; June 20-25, 2009; Miami, FL. [doi: [10.1109/cvpr.2009.5206848](https://doi.org/10.1109/cvpr.2009.5206848)]
65. Kingma D, Ba J. Adam: A method for stochastic optimization. 2015 Presented at: 3rd International Conference on Learning Representations, ICLR; May 7-9, 2015; San Diego, CA.
66. Vahia V. Diagnostic and statistical manual of mental disorders 5: A quick glance. *Indian J Psychiatry* 2013 Jul;55(3):220-223 [FREE Full text] [doi: [10.4103/0019-5545.117131](https://doi.org/10.4103/0019-5545.117131)] [Medline: [24082241](https://pubmed.ncbi.nlm.nih.gov/24082241/)]
67. Pham L, Vu TH, Tran TA. Facial expression recognition using residual masking network. 2021 Presented at: 25th International Conference on Pattern Recognition (ICPR); January 10-15, 2021; Milan, Italy. [doi: [10.1109/icpr48806.2021.9411919](https://doi.org/10.1109/icpr48806.2021.9411919)]
68. Lee JR, Wang L, Wong A. EmotionNet Nano: an efficient deep convolutional neural network design for real-time facial expression recognition. *Front Artif Intell* 2020 Jan 13;3:609673 [FREE Full text] [doi: [10.3389/frai.2020.609673](https://doi.org/10.3389/frai.2020.609673)] [Medline: [33733225](https://pubmed.ncbi.nlm.nih.gov/33733225/)]
69. Han S, Mao H, Dally WJ. Deep compression: Compressing deep neural networks with pruning, trained quantization and Huffman coding. 2016 Presented at: 4th International Conference on Learning Representations, ICLR; May 2-4, 2016; San Juan, Puerto Rico.

70. Loth E, Garrido L, Ahmad J, Watson E, Duff A, Duchaine B. Facial expression recognition as a candidate marker for autism spectrum disorder: how frequent and severe are deficits? *Mol Autism* 2018 Jan 30;9(1):7 [FREE Full text] [doi: [10.1186/s13229-018-0187-7](https://doi.org/10.1186/s13229-018-0187-7)] [Medline: [29423133](https://pubmed.ncbi.nlm.nih.gov/29423133/)]
71. Zhao W, Lu L. Research and development of autism diagnosis information system based on deep convolution neural network and facial expression data. *LHT* 2020 Mar 25;38(4):799-817. [doi: [10.1108/lht-08-2019-0176](https://doi.org/10.1108/lht-08-2019-0176)]
72. Clark TF, Winkielman P, McIntosh DN. Autism and the extraction of emotion from briefly presented facial expressions: stumbling at the first step of empathy. *Emotion* 2008 Dec;8(6):803-809. [doi: [10.1037/a0014124](https://doi.org/10.1037/a0014124)] [Medline: [19102591](https://pubmed.ncbi.nlm.nih.gov/19102591/)]
73. Castelli F. Understanding emotions from standardized facial expressions in autism and normal development. *Autism* 2005 Oct 30;9(4):428-449. [doi: [10.1177/1362361305056082](https://doi.org/10.1177/1362361305056082)] [Medline: [16155058](https://pubmed.ncbi.nlm.nih.gov/16155058/)]
74. Adolphs R, Sears L, Piven J. Abnormal processing of social information from faces in autism. *J Cogn Neurosci* 2001 Feb 15;13(2):232-240. [doi: [10.1162/089892901564289](https://doi.org/10.1162/089892901564289)] [Medline: [11244548](https://pubmed.ncbi.nlm.nih.gov/11244548/)]
75. Capps L, Yirmiya N, Sigman M. Understanding of simple and complex emotions in non-retarded children with autism. *J Child Psychol Psychiatry* 1992 Oct;33(7):1169-1182. [doi: [10.1111/j.1469-7610.1992.tb00936.x](https://doi.org/10.1111/j.1469-7610.1992.tb00936.x)] [Medline: [1400699](https://pubmed.ncbi.nlm.nih.gov/1400699/)]
76. Weeks SJ, Hobson RP. The salience of facial expression for autistic children. *J Child Psychol Psychiatry* 1987 Jan;28(1):137-151. [doi: [10.1111/j.1469-7610.1987.tb00658.x](https://doi.org/10.1111/j.1469-7610.1987.tb00658.x)] [Medline: [3558531](https://pubmed.ncbi.nlm.nih.gov/3558531/)]
77. Hobson RP. The autistic child's appraisal of expressions of emotion. *J Child Psychol Psychiatry* 1986 May;27(3):321-342. [doi: [10.1111/j.1469-7610.1986.tb01836.x](https://doi.org/10.1111/j.1469-7610.1986.tb01836.x)] [Medline: [3733915](https://pubmed.ncbi.nlm.nih.gov/3733915/)]
78. Banire B, Thani D, Makki M, Qaraqe M, Anand K, Connor O. Attention Assessment: Evaluation of Facial Expressions of Children with Autism Spectrum Disorder. In: Antona M, Stephanidis C, editors. *Universal Access in Human-Computer Interaction. Multimodality and Assistive Environments. HCII 2019. Lecture Notes in Computer Science*, vol 11573. Cham, Switzerland: Springer; 2019:32-48.
79. Akmanoglu N. Effectiveness of teaching naming facial expression to children with autism via video modeling. *Educ Sci-Theor Pract* 2015:1. [doi: [10.12738/estp.2015.2.2603](https://doi.org/10.12738/estp.2015.2.2603)]
80. Li M, Li X, Xie L, Liu J, Wang F, Wang Z. Assisted therapeutic system based on reinforcement learning for children with autism. *Comput Assist Surg (Abingdon)* 2019 Oct 14;24(sup2):94-104. [doi: [10.1080/24699322.2019.1649072](https://doi.org/10.1080/24699322.2019.1649072)] [Medline: [31411501](https://pubmed.ncbi.nlm.nih.gov/31411501/)]
81. Lecciso F, Levante A, Fabio RA, Capri T, Leo M, Carcagni P, et al. Emotional expression in children with ASD: a pre-study on a two-group pre-post-test design comparing robot-based and computer-based training. *Front Psychol* 2021 Jul 21;12:678052 [FREE Full text] [doi: [10.3389/fpsyg.2021.678052](https://doi.org/10.3389/fpsyg.2021.678052)] [Medline: [34366997](https://pubmed.ncbi.nlm.nih.gov/34366997/)]
82. Bekele E, Zheng Z, Swanson A, Davidson J, Warren Z, Sarkar N. Virtual Reality-Based Facial Expressions Understanding for Teenagers with Autism. In: Stephanidis C, Antona M, editors. *Universal Access in Human-Computer Interaction. User and Context Diversity. UAHCI 2013. Lecture Notes in Computer Science*, vol 8010. Berlin, Germany: Springer; 2013:454-463.
83. Bauminger N. The facilitation of social-emotional understanding and social interaction in high-functioning children with autism: intervention outcomes. *J Autism Dev Disord* 2002 Aug;32(4):283-298. [doi: [10.1023/a:1016378718278](https://doi.org/10.1023/a:1016378718278)] [Medline: [12199133](https://pubmed.ncbi.nlm.nih.gov/12199133/)]
84. Sengupta J, Kubendran R, Neftci E, Andreou A. High-speed, real-time, spike-based object tracking path prediction on google edge TPU. 2020 Presented at: 2nd IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS); August 31 - September 2, 2020; Genova, Italy. [doi: [10.1109/aicas48895.2020.9073867](https://doi.org/10.1109/aicas48895.2020.9073867)]
85. Voss C, Washington P, Haber N, Kline A, Daniels J, Fazel A, et al. Superpower glass: delivering unobtrusive real-time social cues in wearable systems. 2016 Presented at: ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct; September 12-16, 2016; Heidelberg, Germany. [doi: [10.1145/2968219.2968310](https://doi.org/10.1145/2968219.2968310)]
86. Washington P, Voss C, Kline A, Haber N, Daniels J, Fazel A, et al. SuperpowerGlass: a wearable aid for the at-home therapy of children with autism. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol* 2017 Sep 11;1(3):1-22. [doi: [10.1145/3130977](https://doi.org/10.1145/3130977)]
87. Kline A, Voss C, Washington P, Haber N, Schwartz H, Tariq Q, et al. Superpower Glass. *GetMobile: Mobile Comp. and Comm* 2019 Nov 14;23(2):35-38. [doi: [10.1145/3372300.3372308](https://doi.org/10.1145/3372300.3372308)]
88. Voss C, Schwartz J, Daniels J, Kline A, Haber N, Washington P, et al. Effect of wearable digital intervention for improving socialization in children with autism spectrum disorder: a randomized clinical trial. *JAMA Pediatr* 2019 May 01;173(5):446-454 [FREE Full text] [doi: [10.1001/jamapediatrics.2019.0285](https://doi.org/10.1001/jamapediatrics.2019.0285)] [Medline: [30907929](https://pubmed.ncbi.nlm.nih.gov/30907929/)]
89. Haber N, Voss C, Wall D. Making emotions transparent: Google Glass helps autistic kids understand facial expressions through augmented-reality therapy. *IEEE Spectr* 2020 Apr;57(4):46-52. [doi: [10.1109/mspec.2020.9055973](https://doi.org/10.1109/mspec.2020.9055973)]
90. Daniels J, Schwartz JN, Voss C, Haber N, Fazel A, Kline A, et al. Exploratory study examining the at-home feasibility of a wearable tool for social-affective learning in children with autism. *NPJ Digit Med* 2018 Aug 2;1(1):32 [FREE Full text] [doi: [10.1038/s41746-018-0035-3](https://doi.org/10.1038/s41746-018-0035-3)] [Medline: [31304314](https://pubmed.ncbi.nlm.nih.gov/31304314/)]
91. Washington P, Voss C, Haber N, Tanaka S, Daniels J, Feinstein C, et al. A Wearable Social Interaction Aid for Children with Autism. 2016 Presented at: CHI Conference Extended Abstracts on Human Factors in Computing Systems; May 7-12, 2016; San Jose, CA. [doi: [10.1145/2851581.2892282](https://doi.org/10.1145/2851581.2892282)]

92. Daniels J, Haber N, Voss C, Schwartz J, Tamura S, Fazel A, et al. Feasibility testing of a wearable behavioral aid for social learning in children with autism. *Appl Clin Inform* 2018 Jan 21;9(1):129-140 [[FREE Full text](#)] [doi: [10.1055/s-0038-1626727](https://doi.org/10.1055/s-0038-1626727)] [Medline: [29466819](https://pubmed.ncbi.nlm.nih.gov/29466819/)]

Abbreviations

AKDEF: Averaged Karolinska Directed Emotional Faces
API: application programming interface
ASD: autism spectrum disorder
CAFE: Child Affective Facial Expression
CK+: Extended Cohn-Kanade Dataset
CNN: convolutional neural network
ExpW: Expression in-the-Wild
FePh: Facial Expression Phoenix
FER: facial expression recognition
FER-2013: Face Expression Recognition 2013
JAFFE: Japanese Female Facial Expression
KDEF: Karolinska Directed Emotional Faces
NAS: neural architecture search
NIMH - ChEFS: National Institute of Mental Health Child Emotional Faces Picture Set
RaFD: Radboud Faces Dataset
RCT: randomized controlled trial
SIGF: Stanford Interdisciplinary Graduate Fellowship
Tsinghua-FED: Tsinghua Facial Expression Database

Edited by A Mavragani; submitted 29.05.22; peer-reviewed by B Nieves Soriano, S Badawy, B Li; comments to author 27.06.22; revised version received 01.08.22; accepted 09.08.22; published 21.03.23

Please cite as:

Banerjee A, Mutlu OC, Kline A, Surabhi S, Washington P, Wall DP

Training and Profiling a Pediatric Facial Expression Classifier for Children on Mobile Devices: Machine Learning Study

JMIR Form Res 2023;7:e39917

URL: <https://formative.jmir.org/2023/1/e39917>

doi: [10.2196/39917](https://doi.org/10.2196/39917)

PMID: [35962462](https://pubmed.ncbi.nlm.nih.gov/35962462/)

©Agnik Banerjee, Onur Cezmi Mutlu, Aaron Kline, Saimourya Surabhi, Peter Washington, Dennis Paul Wall. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 21.03.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.