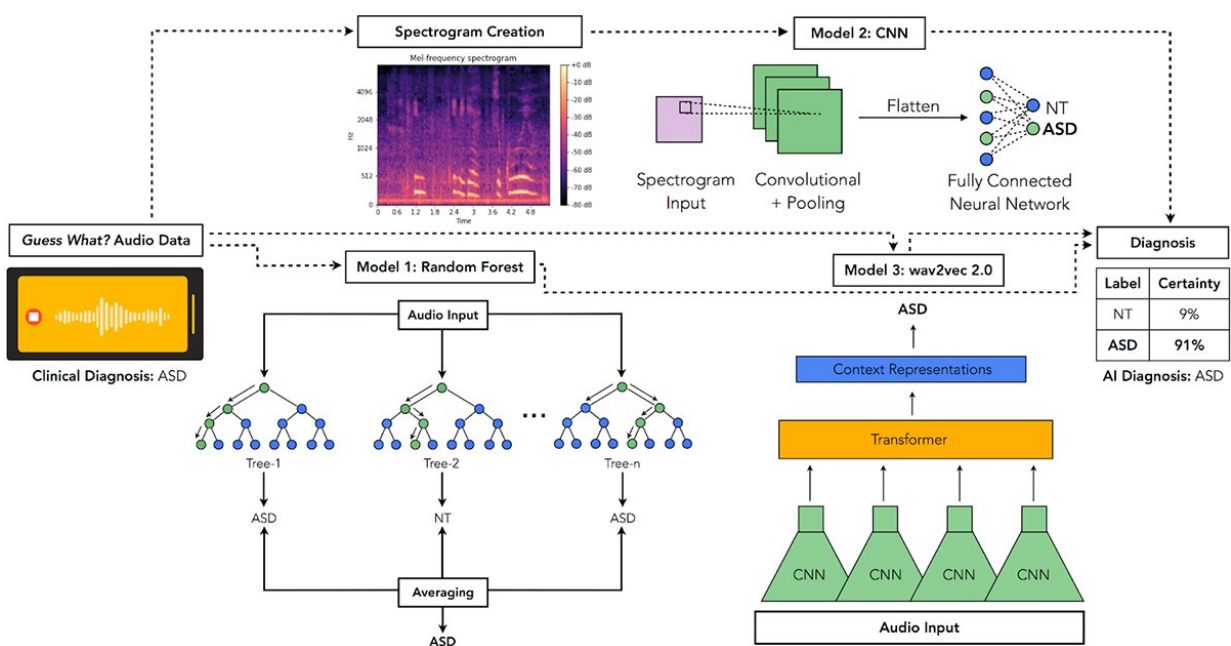# Smartphone game data could facilitate earlier diagnosis of autism spectrum disorder and improve treatment

September 28 2022, by Katharine Miller

Overview of audio-based AI detection pipeline. First, the educational video game Guess What? crowdsources the recording of videos of NT children and children with ASD from consenting participants. Audio of children's speech is manually spliced from the videos and 3 models are trained on this audio data. The first is a random forest classifier, which uses an ensemble of independently trained decision trees. The second is a CNN. The third is a fine-tuned wav2vec 2.0 model. Model 1 takes commonly used speech recognition features as input, model 2 learns from spectrograms of the audio, and model 3 takes the raw audio data itself as input. AI: artificial intelligence; ASD: autism spectrum disorder; CNN: convolutional neural network; NT: neurotypical. Credit: *JMIR Pediatrics*

In the game [Guess What?](), created by Stanford researchers, an adult caregiver holds a smartphone to his or her forehead and asks a child to mimic an image displayed on the screen. It might be a monkey, a soccer player, or perhaps a happy or sad face. The adult then guesses what the child is acting out and registers correct answers by tilting the phone forward; incorrect by tilting it back.

For [children]() with [autism spectrum disorder]() (ASD), the game provides a quick dose of therapeutic learning in the home setting—helping them make [eye contact]() with their caregivers as well as helping them associate specific emotions with various facial expressions.

But the value of Guess What? goes much deeper. Each 90-second game session is video recorded and can be submitted (with appropriate consents and privacy protections) to Stanford researchers.

"If we switch the camera on, and we can give useful prompts to the child, we can challenge them, help them, and capture information as we go," says Dennis Wall, professor of pediatrics, of psychiatry and [behavioral sciences](), and of biomedical data science at Stanford Medicine and an affiliated faculty member with the Stanford Institute for Human-Centered Artificial Intelligence.

For a few years now, Wall and his colleagues have been gathering Guess What? home video recordings and using them to develop new ways to diagnose ASD remotely, improve emotion-recognition datasets, track children's progress recognizing emotions, and ultimately improve ASD treatments.

The work, which uses computer vision as well as other forms of AI, has potential applications for other types of behavioral analysis as well. "The methods that the Wall Lab has developed for autism can enable new general purpose models of human behavior which can be applied to all sorts of conditions, including other developmental delays, mental health conditions, and affective disorders such as schizophrenia," says Peter Washington, who recently completed his Ph.D. in Wall's lab and is now an assistant professor of information and computer science at the University of Hawaii.

The Guess What? origin story

Children with ASD typically struggle with making eye contact and engaging in what's called social-emotional reciprocity—the back and forth of social interaction that requires an understanding of nonverbal cues including, among other things, the recognition of emotion in others' faces. Social reciprocity is best learned when children are quite young, and various strategies for teaching it to children with ASD—using, for example, handheld flash cards—have proven effective but not generalizable or scalable, Wall says.

To address that problem, Wall and his colleagues developed an autism treatment program that uses augmented reality wearables—specifically Google Glass—to provide children with cues about the emotions of people with whom they are interacting. Although the Google Glass approach received attention in the press and has proven effective in a randomized controlled trial, augmented reality tools have not yet been widely adopted, Wall says.

To overcome that limitation, Wall and his team developed Guess What? which relies on a more ubiquitous tool: the smartphone. "Most people across all sectors of socio-economic status, race, and ethnicity have a smartphone," Wall says. "This makes it a potent vehicle for helping to

manage health and deliver treatments." The smartphone is also more natural for families to use, Wall says, because it creates an excuse for social exchange in the home.

But Wall also had another goal in mind: Collecting home videos of both ASD and neurotypical children. By creating a game that makes it easy for families to safely share videos with researchers, they hoped to gather a large enough dataset to move the field of ASD diagnosis and treatment forward. And those efforts are beginning to bear fruit.

Diagnosing autism early and remotely

Researchers know that early intervention is beneficial to children with autism, yet in the United States, diagnosis typically takes about two years and the average age of diagnosis is nearly 4½, Wall says. In addition, autism services are not uniformly distributed around the country: More than 83% of U.S. counties offer no autism diagnosis services at all.

To fill in that gap in service, Wall's team is using Guess What? home videos to develop AI models that could eventually allow semi-automated remote diagnosis of ASD.

And there's good reason to think they could succeed. In earlier work, Wall and Washington found that crowdsourced nonexperts are just as good as clinicians at reviewing videos of youngsters and labeling meaningful features of ASD (such as certain types of repetitive speech, a lack of eye contact, certain head movements, or idiosyncratic use of fingers, such as finger picking). Further: Using these labeled videos they trained a model to predict with high accuracy (above 90%) which children are neurotypical or have ASD in another set of videos (one that has also been manually labeled with meaningful features of ASD).

Going forward, the team hopes their models will gradually become smart

enough to diagnose a [child](link) with ASD without human assistance. To make that leap, however, the team needs a sizable stock of home videos to use for model training. That's where Guess What? comes in.

In a recent paper published in *JMIR Pediatrics and Parenting*, the team tested the idea of using just the audio portion of Guess What? video recordings to directly predict ASD without relying on any humans to label relevant features. Audio is relevant to ASD diagnosis because many children with autism are known to vocalize differently than neurotypical children, Washington says. For example, they often echo back the words used by others, speak in a monotone or at an atypical pitch, and stress their words in unusual ways.

Applying audio-based deep learning methods to 850 audio clips of 58 children (20 with ASD), the team achieved 79% accuracy in distinguishing neurotypical children from those with ASD, Washington says. And while that is not clinically adequate, he says, it's important to keep in mind that because ASD presents in many ways in different people, one wouldn't expect audio recordings alone to suffice for diagnosis.

Indeed, he says, "Even the crowdsourced machine learning models that we developed and that use human-in-the-loop feature extraction would still require at least five different behaviors or input features to achieve high performance."

In next steps, the team will start combining the audio signal with other types of behavioral information in the videos such as emotion recognition, hand movements, etc. "I'm interested in developing multimodal models that integrate multiple sources of data into one explainable diagnostic system," Washington says. "It will be useful to know how much of an automated diagnosis was attributed to, for example, emotion recognition or eye contact versus another behavior

such as speech."

Building better emotion recognition datasets

Not only do the scholars hope their game will help diagnose ASD early, they also hope it will assist children in learning and recognizing others' emotions. When children look at another person's face using the Google Glass autism system developed by Wall and his team, the glass will tell them what emotion the other person is expressing.

For the system to do this accurately and reliably requires a model trained on datasets of labeled facial expressions. In recent work, Wall, Washington, and their colleagues extracted images of children's facial expressions from a large trove of Guess What? videos. They then created a "Hollywood Squares" type game that allowed nonexperts to quickly review and label the images with various emotions. For example, the labelers could mark an entire swath of images as happy or sad rather than having to click on each picture.

The result: A model trained with this labeled dataset performed better than any other existing model at identifying emotions in a standard benchmarked dataset of children's facial expressions.

Opportunities going forward

As the Guess What? dataset grows, it's likely to generate information that can be fed back into the game to improve its effectiveness as a teaching and treatment tool, Wall says.

At the same time, Wall is eager to make use of the team's ASD models to track how children with ASD are faring over time. For example, are they getting better at emotion recognition or not? Are there changes in

their head movements, finger movements, and eye contact or not?

"Combine this information together and you suddenly have the potential for an ongoing longitudinal mechanistic understanding of the progression of the ASD phenotype, which is a complete game changer," he says. "Suddenly you have a set of time series data, all from playing a game."

More information: Nathan A Chi et al, Classifying Autism From Crowdsourced Semistructured Speech Recordings: Machine Learning Model Comparison Study, *JMIR Pediatrics and Parenting* (2022). DOI: 10.2196/35406

Provided by Stanford University