Using Gaze Tracking on Sensory Integration Therapy for Eye-movement Detection

Yirui Yin

Shanghai Starriver Bilingual School

ABSTRACT

Autism spectrum disorder is a condition that often occurs in infancy (from birth to three years old). A common problem includes eyesight disorders, where children cannot focus on a specific item. We use the method of Convolutional Neural Networks (CNN) to detect eyes and collect eye positions. Then, we use a linear regression model to best fit the movement of the eyeballs through the track of positions. After that, we use an Analytic Hierarchy Process (AHP) model for a multi-perspective judgment on children's performance in eye movement.

Background Information

Introduction

Autism spectrum disorder is a developmental disability that severely inhibits the ability to communicate, learn, and interact socially. Symptoms include hypersensitivity or hyposensitivity to stimuli, such as excessive fluorescent or noise, which can last throughout the patient's life. Despite the severity of the disorder, many people are not aware of it due to its high price to be localized in the community. Using AI to identify eye positions, my schoolmates and I (the following paragraphs may use the pronoun “we”) created a product used for sensory integration therapy.

Firstly, we had to deal with image spotting and analysis, which included two objectives: detecting the eyes and reporting the eye positions, so that a trend of eye movement could be predicted. Thus, we used a Convolutional Neural Network (CNN) [1] trained eye detection model. It is an artificial intelligent algorithm that mimics human decision-making. The algorithm can distinguish eye position by making a comparison on how likely the specific image is to the eye. Using the Gaze Tracking package on GitHub [2], we were able to spot the correct position of the eyes. In multiple trials, Yolo-V5 was the best fitting algorithm for object detection. However, there were essential weaknesses, including only one single person could be detected at once and heavily relying on original datasets. In the essay, CNN has been introduced in the mechanism. The CNN has multiple layers under two different layers: convolutional layers and pooling layers.

Secondly, as we had the data from the first step, we had to use a mathematical function to derive the movement of the eyeball. A linear regression model [3] could reduce the residual from the detection and conclude a line with a slope that could be compared to the movement that children follow. The best result is that the eye movement fits the line. The greater the difference, the lower the performance score.

Thirdly, the performance is not reasonable for just one dimension. Different doctors have different points of view on performances. We will need a model for connecting multiple feedback and a uniformed performance score for the children. A vector scale is used to normalize all the data in eye movement results, doctor feedback, parents' feedback, and other aspects, including objective and subjective points of view. As a requirement, we used an Analytic

Hierarchy Process [4] (AHP) model for the overall score, which is a matrix that has a weight on each dimension. In addition, we had to normalize all data to a range of zero to one to make a standardized score range.

Focus

My innovation focuses on solving autism spectrum disorder, which is a kind of neurodevelopmental disability that severely inhibits humans' ability to communicate, learn, and interact socially. It incorporates extreme introverted behavior and Asperger syndrome. People with autism frequently encounter challenges with social communication and interaction and exhibit limited, repetitive patterns of behavior or exercises. Symptoms are classified into two categories: social communication and behavior patterns. For social communication and interaction, it includes poor eye contact and facial expressions, disability in speech, misunderstanding of questions, aggressive or disruptive temper, and repeating words or phrases. For behavior patterns, it includes repetitive movements like hand shaking or spinning, lack of physical movement coordination, high sensitivity to sound or touch, and engaging in self-harming activities. Symptoms are typically found in children between the ages of one and two years old, who show a high disorder inflection rate. Long-term issues may include challenges in performing everyday tasks, creating and maintaining relationships, and keeping a job. Autism is the fastest-growing developmental disability in the world. The known causes for this disorder remain in genetics and environmental factors. For genetics, certain inherited gene mutations and genetic disorders such as Rett syndrome or fragile X syndrome could be the cause of autism disorder. For environmental factors, researchers are examining medication, viral infection, and issues during pregnancy to be the cause of this disorder. In 2020, the CDC [5] reported that approximately 1 in 54 children in the U.S. is diagnosed with an autism spectrum disorder, including a four times higher inflection rate for boys than for girls. Considering the fact that more than 75 million people worldwide suffer from the inability to have regular interaction and the sensitivity toward surrounding stimuli caused by autism, it's crucial to find a solution for those left-behind children with autism, who desperately need affordable and suitable treatments.

For a long time, hospitals have developed a systematic regulation system and administration system for diseases, curing people with serious difficulties. However, sensory integration therapy (SIT) is not equivalent to a disease, but a boost to children's growth, an enhancement rather than a medical treatment. The differences in urgency make it easier for most people to find a hospital than a center for sensory integration therapy. The current access to SIT is time-consuming, expensive, and rare. The innovation is dedicated to solving the pain point in the market.

Medical Background

My innovation is based on Occupational Therapy (OT). It is a complex therapy that includes the training of information processing capabilities of the visual, auditory, olfactory, and tactile senses. Sensory Integration Therapy (SIT) helps develop the corresponding nerves for each of these senses, solves visual irregularities, weak hearing and inability to recognize commands, improves the ability to interpret sound information, command information, and environmental information, and improves sound orientation. It also improves tactile perception by identifying prodding, caressing, pressing, different malleability, stiffness, and temperature.

Occupational therapy also includes the training of balance ability and muscle control, improves balance perception, and improves nerve control. It helps establish static equilibrium in standing and sitting. Occupational therapy includes treatments for dynamic balancing such as cycling and sliding.

Occupational therapy can be used not only as a treatment for autism but also as a treatment for elders who have lost their ability to perform daily activities independently due to disease or old age. The therapy mainly focuses on neuron rehabilitation; thus, this innovation does not provide a solution for the recovery of autism. This innovation can also be used by normal kids, helping them to develop neurons more efficiently and thoroughly. The broad potential usage of the innovation shows its promising future.

CAD Graph

Composition

The composition is a combination of three main parts and six specific products. The three main parts are the LCD screen, BOSU ball, and an air cushion. The LCD screen also includes four specific products, including the stereo (speaker), projectiles, LCD screen, and camera.

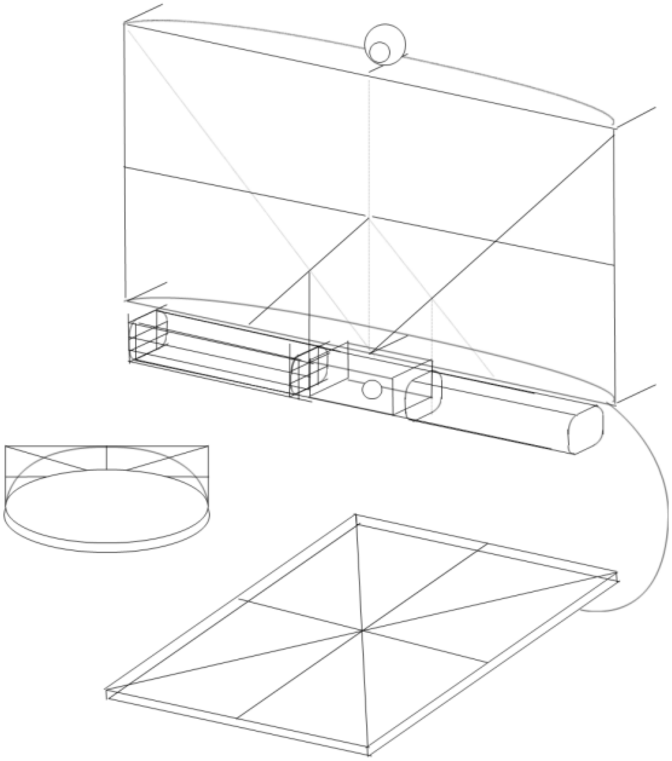

Figure 1. A general overview

Figure 1 introduces our product, which is designed with three main parts. The upper part is an LCD screen with a camera and two speakers on each side. The screen serves as direct communication between the algorithm and the patients. The camera will take a video of the entire therapy session and use the algorithm introduced in the following essay to determine the performance results based on the eye movements.

The bottom left part is a BOSU ball that children can interact with. Tactile feedback is crucial for sensory treatment. The BOSU ball can respond to the pressure that children apply to it and the force they exert on the ball. With a systematic evaluation, a mathematical model for performance results could be developed. It is unfortunate that no such experiments have been conducted to build a mathematical model for calculating the results of tactile performance, which is an area that future researchers could focus on.

The bottom right part is an air cushion that can detect the movement of the children. It can appeal to children and increase their interest in playing, interacting, and receiving therapy.

Detection

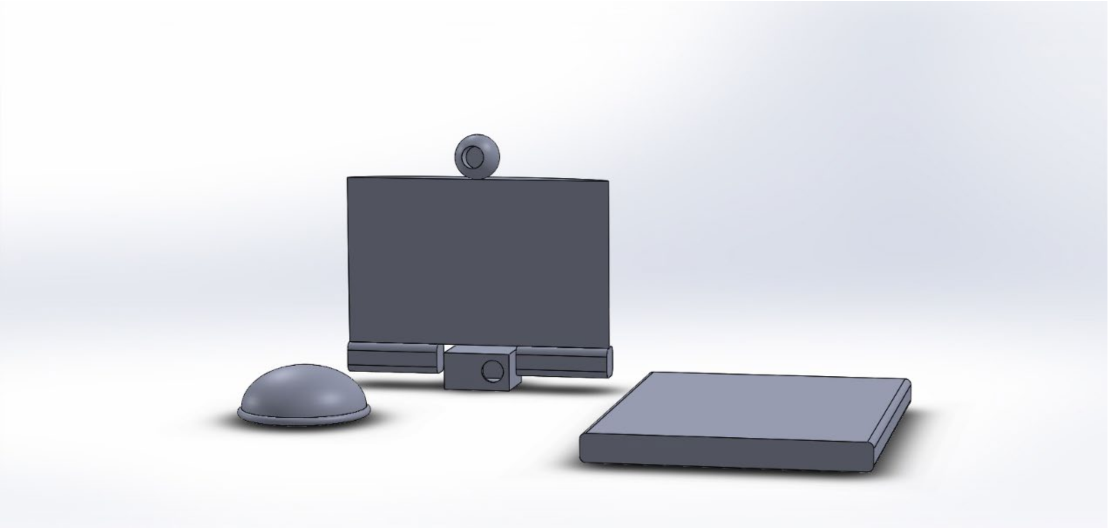

Figure 2. The 3D images

The figure 2 provides a picture of all three components. The detection occurs on the LCD screen with the camera on top of it. The projector at the bottom saves space for a thinner screen and reduces construction costs. Additionally, it is easier to fix instead of changing every part in case of a mechanical error.

The camera can take videos and connect with computers through the internet. The algorithm can run on computers and phones, which helps to send the results, performance scores, and video records to professional clinics for further evaluations.

Methods

Assumption

Assumption 1: All data are valid, credible and realistic value under normal status.
Justification 1: It is a risky assumption. The application and data are not under fully tests and a big range of random samples. However, as an explanation and justifying the method. We will assume the data are all right, including a possible false in the data that we collected.
Assumption 2: We do not assume the children will be distracted. The participants pay full attention on test.
Justification 2: In real life, I have tested on young children who are close relatives to me and my friends. All children are engaged with tests and trainings. Although it is too ideal that every child will participate with trust on machine, we

need to eliminate the possible influence to the data that we cannot control.

Convolutional Neural Networks

Convolutional Neural Networks, or CNN in short, is an algorithm that close to neural networks. CNN performance well on objects detection which is a boost to the eyeball detection. The method contains weights and biases, with RELU to eliminate some of the blocks on the pictures. There are a lot of details in the object detection methods which the essay cannot cover in short paragraphs. In this section, instead of demonstrating how a picture works in CNN, we will use a $3 * 3$ matrix, which simulates the width and height of a picture with 9 pixels. If it is a picture with color that fills 8 digits on computer and as big as a HD size, the vector should be $1280 * 720 * 256$, which is too complicated to explain.

Convolutional Layer

The convolutional layer is a layer that uses different filters to converge a large picture into a smaller vector in size. The filter depends on the depth of the image, which is related to the color of the picture. The algorithm generates learnable "neurons" through filters.

However, for a large picture, there are many places that are not taken into consideration. Even for eye detection, people use an image of the half-body for testing. Therefore, there will be many neurons, making it impossible to make a full connection at every level. Thus, we connect neurons to a part of the input volume. The spatial extent of this connectivity is a hyperparameter called the receptive field of the neuron. In figure 3, as an example, the left picture demonstrates a connection to a local region. The five outputs have different weights but share the same receptive field. Each output is generated by a distinguished filter, giving them a different value and weight. The right part of figure 3 provides a mathematical explanation for the process.

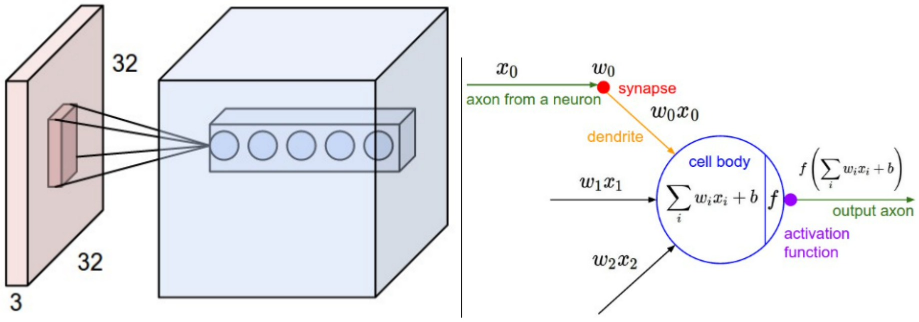

Figure 3. A figure that explains the function and process of building CNN in image [6]

If we are not taking stride into the count, the number of neuros could be expressed as the following with parameters on table 1:

Table 1. Parameter and Symbols in equation

Parameter	symbols
input volume size	W
receptive field size of the Conv Layer neurons	F

amount of zero padding used on the border	P

The mathematical expression for the number of neurons is the following:

Equation 1: the number of neurons calculated by the variables in the table 1.

$$number\ of\ neurons = (W - F + 2P) + 1$$

Using a matrix for explanation, the following matrix is the input volume, with a two square receptive field:

$$\begin{bmatrix} 1 & 0 & 2 \\ 1 & 1 & 0 \\ 2 & 0 & 0 \end{bmatrix}$$

The filter is the following matrix:

$$\begin{bmatrix} 1 & 0 \\ -1 & 0 \end{bmatrix}$$

The output will be:

$$\begin{bmatrix} 1+1+1+0+1 & 0+2+1+0+0 \\ 1+1+2+0-1 & 1+0+0+0+0 \end{bmatrix}$$

Or

$$\begin{bmatrix} 4 & 3 \\ 2 & 1 \end{bmatrix}$$

Pooling Layer

The convolutional layer reduces the size of the original matrix. The neural network is trained through repeated trials. Therefore, we need to expand the matrix back to its original size using the following steps.

1. The results after Conv Layer $W * H * D$
2. Two experimental data: spatial extent and stride F, S
3. The new size is calculated by the following expression, including an unchanged depth:

$$W_{new} = \frac{W - F}{S} + 1$$

$$H_{new} = \frac{H - F}{S} + 1$$

4. Fills the edges with 0 if the new width and height is not equal to the original expression.

Linear Regression

We generated a linear regression model based on eye movements, attempting to fit the linear moving dots in the animation. In real life, the movement is close to linear regression, but it never fits a straight line perfectly. People are animals that move with some randomness, unlike machines. Therefore, there will be random errors and residuals in linear regression.

According to basic statistical theory, the trend of a sample is not equivalent to the trend of the population. Using the sample, a unary linear regression line can fit the trend. Since all functions and fit lines are written in linear form, ordinary least squares (OLS) can be used to find the least residual errors, which is the fundamental principle for all regression methods.

To simplify the model, we use a matrix and the linear model to express the relationship. The linear regression is expressed as follows:

$$Y = a_0 + a_1 X + e$$

where a_0 and a_1 are the coefficients. However, for most of the cases in statistics, there is no absolute coefficient. In our case, the machine could lose accuracy under problems such as low light, low battery, broken camera. As a result, it is impossible to find the exact coefficient. So, the model is looking for the best fit a_0, a_1 with random error e:

$$E(e) = 0, \qquad 0 < Var(e) = \sigma^2 < \infty$$

Ideally, the residuals are normally distributed around the best-fit line. This is a reasonable simplification for the model. In real life sense, the y is equivalent to the position of eyes on the y axis. The x in the equation is the position of eyes on the x axis. While a_1 is the velocity of the eye's movements.

According to the methods of OLS, the sum of all residual's square is the following:

$$\min Q(a_0, a_1) = \sum_{i=1}^{n} (Y_i - \hat{Y})^2 = \sum_{i=1}^{n} [Y_i - a_0 - a_1 X_i]^2$$

For the multivariable function Q, the lowest value could be determined by solving the function:

$$\begin{cases} \dfrac{\partial Q}{\partial a_0} = -2\sum_{i=1}^{n} [Y_i - a_0 - a_1 X_i]^2 = 0 \\ \dfrac{\partial Q}{\partial a_1} = -2\sum_{i=1}^{n} X_i [Y_i - a_0 - a_1 X_i] = 0 \end{cases}$$

These two equations are the best possible solutions to use for the estimation. All models are approaching a minimal Q value representing a low residual: the actual points deviate less from the estimation points. This method is the fundamental reason why other estimations work out.

Thus, the standard equations for linear regression are as the following:

$$\begin{cases} m a_0 + (\sum_{i=1}^{m} x_i) a_1 = \sum_{i=1}^{m} y_i \\ (\sum_{i=1}^{m} x_i) a_0 + (\sum_{i=1}^{m} x_i^2) a_1 = \sum_{i=1}^{m} x_i y_i \end{cases}$$

There are significant benefits to this model. On this page, the method is simplified into a first-degree linear model. It could be raised to different degrees if needed. Similar ideas, such as finding the least residual, are also applied to the higher degrees. Most x and y in equations can be replaced by matrixes with more variables. Using SPSS, we conclude the result into the following prediction line:

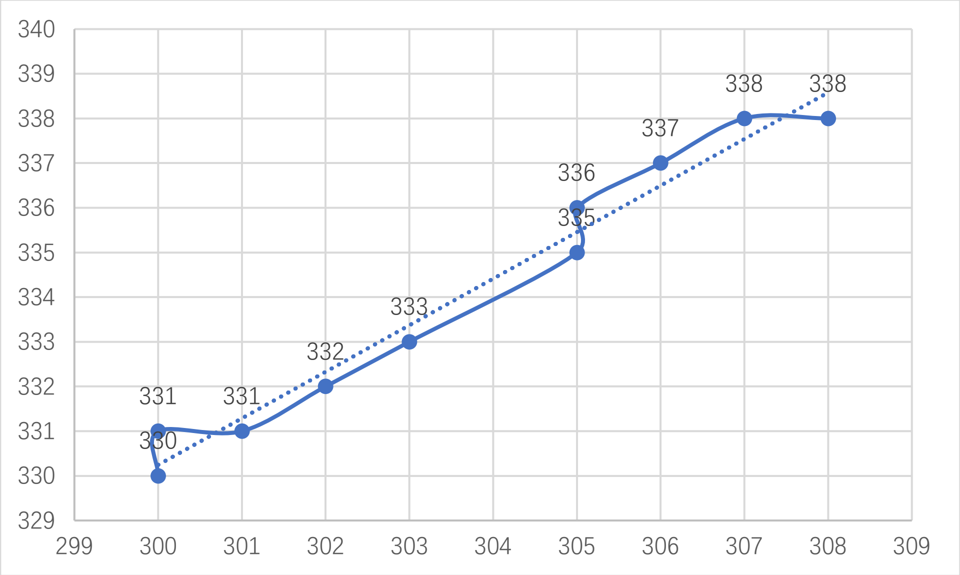

Figure 4. Linear regression for eye position with a best fit line in single variable function

In the table below, we conclude the equation and residual:

Table 2: Main equation and residual

Name	Expression
Equation	y = 1.042x + 17.629
R²	0.9733

Rating System

Data Processing

To avoid the effect of different values having different magnitudes and units, it is necessary to normalize each data point. We must find the maximum and minimum values within the range and transform all data to a zero-to-one scale. By doing so, we can create an objective score.

$$x_i = \frac{x_i - x_{min}}{x_{max} - x_{min}}$$

Applying all the data, we can make a AHP model in 3.4.2.

Weighting Model

As doctors may judge patients differently, individuals may have varying focuses in sensory integration therapy with different weights on visual, tactile, and olfactory modalities. Therefore, we require a rating method to provide a relatively objective score for reference in urgent situations. This method involves a systematic analysis, both qualitative and quantitative, to evaluate the problem.

We define a_{ij} as the result of comparing the i index to the j index. The result should be reference to the table below.

Table 3: The 1-9 scaled table

Scale	Meaning
1	i *and* j are equally important
3	i is slightly important comparing to j
5	i is fairly important comparing to j
7	i is strongly important comparing to j
9	i is very essential comparing to j
2,4,6,8	Mid-value between scales

We can define the results of three different sensory modalities -- visual, tactile, and auditory -- into one vector, taking into account their relative importance to one another. This yields a final measure of the patient's performance.

Strength and Weakness

Strength

The current treatment for autism spectrum disorder heavily relies on private clinics, special treatment centers, or children's care centers, all of which require a significant investment of time, money, and effort. Our product is a small, mobile machine that takes up only one cubic meter of space. Compared to traditional treatments which require a variety of materials, our product saves at least 80\% of space and around 50\% of the cost. The data collected by the machine can be compared across different clinics, and it saves people time depending on their location.

The machine can cure many cases and provide healthcare services along with professional data for doctors. According to experts, eight out of every ten people with autism spectrum disorder suffer from sight problems, and losing direction often accompanies auditory disabilities. The machine focuses on training the three most critical perspectives in sensory integration therapy while maintaining a portable size and reasonable cost with accessible materials.

Doctors can get a general view of children's behavior without requiring long-term observation by examining the scale of results. This revolutionary medical support, coupled with data analysis skills, can save a great deal of time. One significant problem in sensory integration therapy is that people use subjective models rather than objective checklists for treatment. We hope that the machine can lead to a trend of making usable measurements for diagnosing children. The same idea could apply to elderly people with fatal diseases and rare, time-consuming diseases. The impact is not only on saving time for patients and doctors but also on opening a trend of data collection from home and digital healthcare services. If investors can foresee a useful application in children's therapy, they will have much confidence in future medical revolutions that could save millions of lives.

Weakness

There are few weaknesses that weaken the results.

Finding a best fit model for eye detections is hard. Different machine learning models can generate different results. Moreover, the data are meaningless for most of the times. Most dataset on websites are adults' dataset, which will significantly reduce the accuracy when the applications are worked on children. The datasets are rare to find on ASD or SIT.

Secondly, the data are just assistant, but the curing is still on a very subjective point of view. In real life circumstances, the webcam may lose the testing ability, producing position lack of trusts. We don't think doctor will really trust the result. However, there is a time requirement for understanding and adaptation. In the earlier stage, people could use the machine for a supplement or reference.

Thirdly, the application is not under massive tests. The application is still a project under fully tests. We don't know if there are significant problems in real life application. Moreover, we haven't done test on children who really need sensory integration therapy, yet. Testing on patients are really risky. We made all the test to prove that the machine that theoretically works out for most cases.

Discussion

The model has been tested on designers, and all designers and their friends found the dot on the screen easy to follow, with an average score of 0.93. Some designers have also asked their younger siblings to engage with the app, and the results show that young children can easily follow the dot with a score of 0.97. This could be due to children's eagerness to participate in games. If this assumption is reasonable, the game-like therapy product could be a lovely toy for children instead of a cold medical machine. Although there is a slight reduction in data collection efficiency for children, it is a feasible problem to solve. The eyes are harder to detect in children because datasets lack examples from children. We tried with older age groups and different genders as tests, and similar declines in detection efficiency occur in older age groups, but at a higher rate than in children. We believe children's eyes are much different from adults for the CNN model in eye detection.

We have had multiple conversations with experts on children's development and people working in the online medical care system, and they have come to a similar conclusion that online treatment and remote diagnosis are the future trend for solving diseases, especially for rare diseases known by only a minority of doctors. Remote diagnosis can significantly reduce costs and time. However, there are weaknesses, including a lack of data, unclear cameras, and other unaffordable risks. Without a doubt, people trust themselves more than machines. We believe a combination of in-home therapy and in-clinic diagnosis can help reduce risks and problems. The camera is clear enough to record all behaviors and match them with specific data. As a result, users can easily choose the right clips to make judgments on how the children behave

References

Simonyan, K., \& Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. In Proceedings of *the International Conference on Learning Representations (ICLR 2015)*. Retrieved from https://arxiv.org/abs/1409.1556

Antoinelame. “Antoinelame/Gazetracking:Eye Tracking Library Easily Implementable to Your Projects.” *Github*, https://github.com/antoinelame/GazeTracking.

James, G., Witten, D., Hastie, T., \& Tibshirani, R. (2021). An Introduction to Statistical Learning: with Applications in R (Springer Texts in Statistics) (2nd ed. 2021). Springer.

Saaty, Thomas L. "Decision making with the analytic hierarchy process." *International Journal of Services Sciences*, vol. 1, no. 1, 2008, pp. 83-98. doi: 10.1504/IJSSCI.2008.017590.

“Data \& amp; Statistics on Autism Spectrum Disorder.” *Centers for Disease Control and Prevention*, Centers for Disease Control and Prevention, 4 Apr. 2023, https://www.cdc.gov/ncbddd/autism/data.html.

Gopi, Ajay. “EECS at UC Berkeley.” *To Send or to Not Send: A Case Study on Computer Vision for Low Power Edge Devices*, EECS at UC Berkeley, May n.d., https://www2.eecs.berkeley.edu/Pubs/TechRpts/2020/EECS-2020-90.html.