# AI-based Web Application for Children with Autism Spectrum Disorder

Nayeong Choi[2014310316], Jongwoo Lee[2016311266], Seyeon Kim[2018312979], Yongtaek Kim[2018313912], and Jiwon Kim[2018314620]

Sungkyunkwan University
2066, Seobu-ro, Jangan-gu, Suwon-si, Gyeonggi-do, Republic of Korea
dudskchl@skku.edu, jjong8256@g.skku.edu, seyeonkk99@g.skku.edu, kyt5343@g.skku.edu, kjiwon1999@g.skku.edu

**Abstract.** Autism spectrum disorder has recently become increasingly popular with people as they struggle with their ability to adapt and interact. This project aims to help children with autism spectrum disorders learn language, emotion and sensory functions through the development of web applications using AI computer vision technology. This project applies ABA therapy and PECS programs, and provides functions such as '감정놀이방', '행동놀이방', and '칭찬나무방'. In addition, object recognition is performed using YOLOv8, the latest deep learning model, and real-time webcams provide experiences for autistic users to execute and achieve on their own. The goal of this project will help improve the learning environment and quality of life of autistic users by overcoming the limitations of existing educational programs and developing web applications with high accessibility and effectiveness.

**Keywords:** ADS · AI · CV · YOLOv8.

## 1 Introduction

Last year, the public's interest in autism was increased by the drama 'Weird Lawyer Woo Young-woo' about lawyers with autism spectrum. But in reality, the plight of people with autism and their families has not changed much. Autism Spectrum Disorders (ASD) is a developmental disorder that affects the way in which people communicate, interact, and perceive with their surrounding communities. Featuring a wide range of symptoms and severity, one of the major challenges faced by children with ASD is learning sensory functions that are important for language, emotion, and social adaptation and interaction. Existing educational programs for autistic children are somewhat ineffective and accessible, and cost-effective. As a result, it highlights the urgent need for effective and innovative interventions and support systems to help children with ASD overcome these challenges.

In this project, we aim to develop a learning program that utilizes AI computer vision technology to help users with ASD learn language, emotion, and sensory functions more effectively and accessibly. The goal is to overcome the

limitations of traditional teaching methods by integrating existing ASD therapy programs, Application Behavior Analysis (ABA) and Picture Exchange Communication System (PECS) programs, and to provide more attractive and effective learning experiences for ASD users.

The learning program consists of '감정놀이방', '행동놀이방' for learning, and '칭찬나무방' for rewards. '감정놀이방' and '행동놀이방' use YOLOv8, the latest machine learning model for facial expression and object recognition, along with real-time webcams for immersion and realism in ASD users. After taking quizzes about facial expressions and objects, evaluate whether users performed them correctly by showing the expressions or objects directly to the webcam. '칭찬나무방' is a function to improve motivation for ASD users with favorite character songs, videos, or real snacks through praise stickers. This approach not only improves realism, but also allows autistic users to perform and achieve tasks on their own, improving overall learning effects through achievement and motivation.

In addition to the AI model, we tried to develop a web application-type program using tools such as HTML, CSS, and JS in the front-end part and Python, Flask, Ngrok, and GCP in the back-end part. This approach can be expected to significantly improve accessibility and reduce care for carers.

## 2 Design

### 2.1 Architecture

The overall development process is shown in Figure.1. First, when a user signs up or logs in from the login page, the main page appears. On the main page, there are '감정놀이방', '행동놀이방', '칭찬나무방', '노래보상방', '간식보상방', '감정놀이방' and '행동놀이방' are ASD treatment programs using facial expressions and object recognition using AI models. After taking quizzes about facial expressions and objects, evaluate whether the user performed them correctly by showing the expressions or objects directly to the webcam. '칭찬나무방' provides the ability to get compliment stickers through learning, and user can exchange compliment stickers and rewards for his or her favorite character songs, videos, or real snacks through '노래보상방' and '간식보상방'.

### 2.2 Core Skill

#### 2.2.1 AI model

For the AI model for face and object detection, we chose the YOLOv8 model. YOLOv8 is a new SOTA model that was introduced in 2023 by improving YOLOv5. The SOTA model stands for 'State-of-the-art', which means the most accurate model at the current level, with the 'best results at the moment'.

YOLO has three representative features. The first feature of YOLO is to look at the entire image only once. Prior to YOLO, R-CNN split the image into several sheets and analyzed the image using a CNN model. Therefore, even if Object

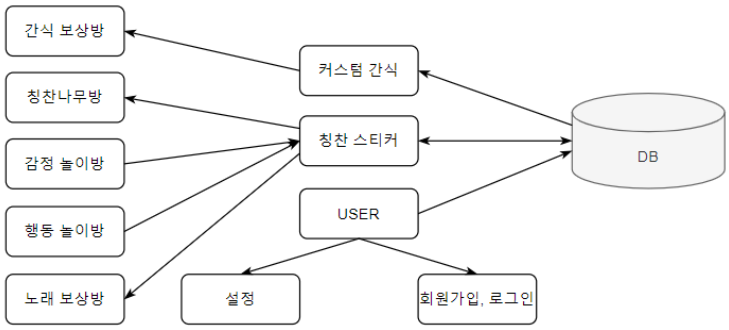


**Fig. 1.** Overall architecture

Detection was performed on one image, it was actually like analyzing several images. However, YOLO has a strong feature of seeing images only once without this process. The second feature is the use of an integrated model. Existing Object Detecting models used a combination of various preprocessing models and artificial neural networks. However, YOLO is simple using an integrated model. The third feature is the ability to detect objects in real time. Real-time webcams can be used to detect objects with performance six times faster than conventional Fast R-CNN. In addition, the model and data format of YOLOv8 are similar to previous versions of YOLO, so it is easy for users to use.

### 2.2.2 Front-end

The front-end was implemented using basic HTML, CSS, and JavaScript files, without the use of specific frameworks. For conventional pages like login, sign up, and main page, HTML and CSS were used to construct the basic structure and design. The main color #007BFF was chosen to maintain a uniform design.

The key pages in this project are the '감정놀이방' and '행동놀이방', which extensively utilize the YOLO model. The '감정놀이방' provides the following features: Webcam Utilization: The navigator.mediaDevices.getUserMedia( video: true ) API in JavaScript was used to get a real-time video stream from the user's webcam. This stream is used to recognize the user's expression and analyze emotions using the YOLO model. Random Display of Emotions and Characters: Various emotions (happy, sad, anger, fear, surprise) and characters within each emotion group are randomly selected and displayed on the page. This was implemented using the Math.random() function in JavaScript. Use of Video and Audio: Video can effectively express dynamic emotions and provide rich information by showing the transition of emotions. The video, which auto-plays when the page loads, is set to muted and loop to play continuously. Audio was added to provide additional information to the user.

The '행동놀이방' has a similar structure to the '감정놀이방' but uses images instead of videos and provides more options.

### 2.2.3 Back-end

In the back-end web implementation, JavaScript (JS) and PHP were utilized, along with MySQL as the chosen DBMS for establishing the database connection. The back-end system was deployed on an GCP VM instance.

To set up the environment, the necessary PHP and Apache2 configurations were set on the VM instance. The web-related frontend and backend source codes were placed in the '/var/www/html' directory. This directory serves as the web root where the application's files are stored. By using JavaScript and PHP together, the back-end handles various aspects of request handling, routing, and processing. JS can handle client-side interactions and enhance the user interface, while PHP handles server-side logic and data processing. The front-end makes requests to the back-end that responds with dynamic content or performs actions based on the requested functionality. The PHP backend communicates with the MySQL database for data storage and retrieval. The backend uses MySQL queries to interact with the database, allowing CRUD (Create, Read, Update, Delete).

In the back-end application implementation, the Python application uses Flask, a popular web development framework in Python, to handle HTTP requests and responses, render HTML templates, and serve static files. The application utilizes Flask-SocketIO, an extension for Flask, to enable real-time bidirectional communication between the server and clients using WebSocket protocol. It demonstrates the usage of event handling and message broadcasting through sockets. The application integrates the Ultralytics YOLO model, a computer vision model for object detection, to perform real-time object detection on video frames captured from a webcam. It showcases the ability to process and analyze video streams using computer vision algorithms.

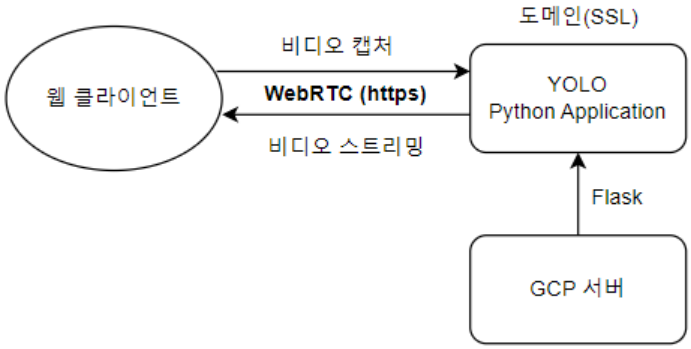


**Fig. 2.** Back-end architecture

## 2.3 Reasoning

### 2.3.1 AI

The most important technology to work on this project is, by far, It is a computer vision technology that recognizes the emotions that appear on the faces of autistic children. Until last year, YOLOv5 dominated the market in object detection, but the game changed with the launch of YOLOv8 in January 2023. Based on the architecture of YOLOv5, YOLOv8 is a SOTA model that introduces higher performance and new improvements, providing additional instance partitioning and tracking capabilities that have never existed before. The reason why our team chose YOLOv8 as the AI model is as follows.

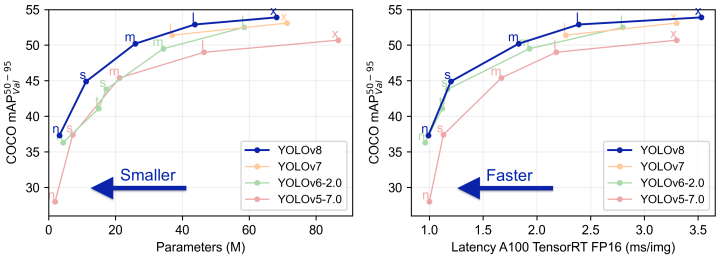


**Fig. 3.** Performance comparison by YOLO version

First, as shown in the Figure.2, YOLOv8 achieves strong accuracy in COCO. For example, the intermediate model, the YOLOv8m model, achieves 50.2% mAP when measured in COCO. It is about 5% higher than the existing YOLOv5m model. As the project focuses on emotion recognition education for autistic children, In addition to increasing the true positive level on the confusion matrix, it was also necessary to consider making the false positive level as low as possible. So our team chose Yolov8 as the AI model with improved mAP performance than Yolov5.

Also, in our team, which has a few people who have never used computer vision, YOLO, which is easier to learn and more convenient than other computer vision tools, was a very good choice. The YOLOv8 comes with a CLI that makes it more intuitive to train models, making it easier to develop. This is added to the Python package, which provides a smoother coding experience than the previous model. Because of this general convenience, even team members who are new to yolo could easily apply it to the program.

### 2.3.2 Front-end

The design choices of the project were primarily focused on creating a child-friendly and intuitive user interface (UI). Particularly, as the primary objective was to build a website tailored for children with autism, it was imperative that

the design remained as simple and intuitive as possible. To this end, we used the basics of HTML, CSS, and JavaScript, without the involvement of any specific frameworks. For these reasons, we endeavored to design a user interface that user could easily understand and use, by minimizing complex components and utilizing larger icons, concise text, and bright colors. We selected the color #007BFF as the main color to maintain a consistent design, notably because it also serves as the symbolic color of World Autism Awareness Day. Moreover, we considered the provision of audio files. Recognizing that some users with autism may have difficulties reading text, we provided audio guidance for these users. This decision was made to increase the accessibility of the design and to ensure that the users could effectively learn through our website.

#### 2.3.3 Back-end

In terms of server deployment, GCP web hosting was chosen as the preferred solution. GCP (Google Cloud Platform) offers a robust and scalable infrastructure that can be configured to meet specific needs. By utilizing GCP web hosting, we benefit from the stability, security, and performance optimization provided by Google's cloud service.

With GCP web hosting, you can configure and deploy independent servers tailored to your application's requirements. This ensures a reliable and secure hosting environment for your backend system. GCP also offers additional features and services that can enhance the performance and scalability of your application, providing a solid foundation for your backend infrastructure. And lastly, MySQL, being a reliable and widely-used relational database management system, was chosen due to its ability to handle complex data structures and ensure data integrity.

In summary, the choice to use MySQL as the database management system and GCP web hosting for server deployment was driven by the need for a reliable and flexible database solution with a clearly defined schema, ensuring data integrity. The decision to utilize GCP's web hosting services was based on the platform's ability to provide a stable, secure, and customizable environment for hosting the backend application. For making a real-time object detection application by YOLO on a web server, we used Flask and WebRTC. Flask is a lightweight and flexible web framework in Python. It provides a simple and intuitive way to handle HTTP requests and responses, making it suitable for building web applications. Flask's modular design allows for easy integration with other libraries and tools, making it a popular choice for web development in Python. WebRTC is a powerful open-source technology that enables real-time communication between web browsers. It provides the necessary protocols and APIs for establishing peer-to-peer connections, allowing for high-quality video, and data streaming directly between browsers without the need for additional plugins or software.

### 2.4 Challenges

#### 2.4.1 AI

The most important challenge in the AI part is increasing accuracy of prediction of the model. To improve the accuracy of the model, we did data verification and augmentation to improve the quality of the dataset. And we set the number of raw-data for each class to be similar. Especially for the '감정놀이방' dataset, we collected dataset with various races. Also we did a lot of tests for various epochs, batch size, image size to find the best model.

#### 2.4.2 Front-end

In the process of front-end development of this project, we encountered two major challenges. The first challenge was considering a child-friendly web design. We tried to minimize the components of the website to make it easier for users with autism to understand. Also, considering that users with autism may need more effort to read characters than other users, we focused on designing the page with immediate experience elements. In this process, we tried to consider from the user's perspective who would use the web page. The second challenge was in providing audio files. In the '감정놀이방', we present videos of virtual characters making various expressions, and the user tries to infer the emotion and solve the problem by watching them. We tried to add audio files to help users who have difficulty in reading or recognizing emotions. However, there were restrictions on browser policy to automatically play audio files when the page is opened. To resolve this, we separately placed a sound play button, which is provided as a symbolic icon for easy understanding. When this button is clicked, three situation descriptions suitable for each emotion are randomly played for each expression of the character.

#### 2.4.3 Back-end

When creating a backend app, the biggest challenge was to create an app that could handle video streaming and to make it work on a web server as well. First of all, we implemented the version that runs locally as the initial form of the app. To do this, we configured an environment where we could handle YOLO and processed videos through opencv. However, when we wanted to run this app on a web server, most of the code that used to work locally didn't work, so we had to find an alternative. WebRTC was used after several trials and errors to allow videos that could be obtained directly from the web server in real time, and for this, it had to support https connection to the web server and encrypt SSL.

## 3 Implementation

### 3.1 AI model (YOLOv8)

As an implementation of the core skill of our AI Model, YOLOv8, YOLOv8 has

three representative features of its architecture. First, YOLOv8 uses anchor free detection. It makes object detection flexible and efficient depending on the situation. It uses Non-Maximum Suppression(NMS). NMS means reducing the number of box predictions, and it can make processing speed of candidates screening after calculation of probability faster. Second, YOLOv8 uses new convolutional networks such as Darknet-53 for a new backbone network, more efficient convolutional neural network, and larger feature map. These features can make the model detect objects more accurately and faster, and capture more complex relationships, and better recognize patterns and objects, and reduce the amount of time for training, and reduce overfitting. Finally, YOLOv8 uses mosaic augmentation. It means the model can combine training images randomly by itself, so that it makes images and objects smaller. And this can help to overcome YOLO's shortcomings, making it difficult to detect small objects. In addition, it makes the effect of learning four images with one image so that it can learn with a small batch size.

### 3.2 Data set

#### 3.2.1 감정놀이방 Data set

On the first screen of '감정놀이방' a video of the person and an audio file depicting the person's situation are presented at the same time. User will work to match the person's facial expressions through video and narration by clicking the button. And after clicking the button, in the webcam screen on the right side of the switched page, the user will conduct work training to copy the expressions just presented.

Our team used a tool called meta human creator provided by unreal engine to provide the appearance of the person. Most of the basic character sets provided by the meta human creator are Westerners, so we created virtual characters as similar to Asian extroverts as possible. Due to the emotional nature of autistic children, they tend to be afraid of new environments and characters, so the figure was limited to three children and four adults. And all seven of the figures are set up to be in the same kindergarten, and the audio file also depicts an event in which each of the characters interact organically.

In the Figure.4, in the order from left to right, are fictional characters named '민지', '민석', '서연', '햇님반선생님', '튼튼반선생님', '달님반선생님', and '원장선생님', respectively.

In addition, the audio voice file was produced using Naver Clova dubbing, and the voice was adopted as 'Hajun's voice', a boy's voice. There are five emotional classes used in the '감정놀이방': joy, sadness, surprise, fear, and anger. Therefore, the audio voice file describing the situation of the person is made into a total of 105 cases, consisting of 7 characters, 5 emotions, and 3 items. The program was organized so that different situations were presented at random each time the learning was executed.

Moreover, an emotion dataset model was created on its own to identify emotions through facial expressions of user on a webcam screen. The model consists of seven classes: positive, fear, suppressed, sad, neutral, angry, and normal, but

**Fig. 4.** Virtual characters created by meta human creator

only five classes are currently used, except for neutral and normal. There are a total of 4879 raw data sets in the emotional dataset, and augmentation work was carried out during model training through YOLO.

#### 3.2.2 행동놀이방 Data set

We made our own dataset for '행동놀이방'. First we downloaded coco dataset for some classes. And we converted other types of dataset on the web to coco type. After combining them, we created and added some short-numbered dataset by using LabelImg program. As a result, we used about 40,000 images of 20 classes, fruits and supplies, for '행동놀이방' dataset model training.

### 3.3 YOLOv8 hyperparameters

The value of the emotion dataset parameter was set to epoch=100, pattern=50, batch=16, imgsz=640. And the value of the object dataset parameter is set to epoch=155, patience=50, batch=64, and imgsz=640. Other parameter values are the same as YOLOv8's default parameter value.

### 3.4 UX/UI viewpoints

The web pages implemented in this project consist of a log-in and sign-up screen, an '감정놀이방' and '행동놀이방', and a '칭찬나무', '노래보상방', and '간식보상방'. Each screen uses sky blue (#007BFF) as the point color. We aimed and implemented an intuitive design so that users can easily understand and use it.

### 3.5 Front-end detail

#### 3.5.1 Front-end structure

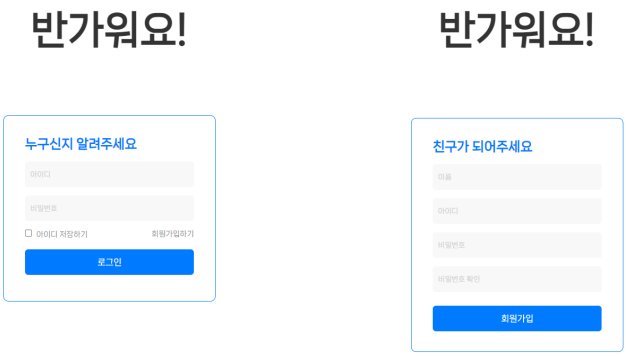


**Fig. 5.** UI example of the log-in and sign-up page

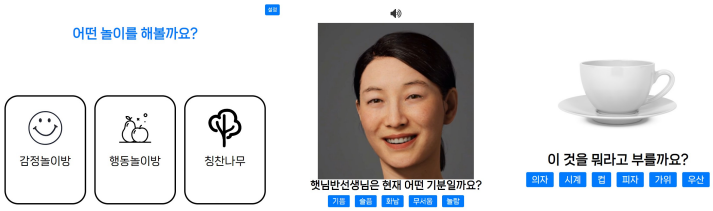


**Fig. 6.** UI example of the main page, 감정놀이방 and 행동놀이방

The front-end of our application begins with a login screen, which also offers a path to a sign-up screen. On successful sign-up, the user is redirected back to the login screen where they can use their newly created credentials to log in. Upon successful login, users are directed to the main.html page which connects four different sections of our application: '감정놀이방', '행동놀이방', '칭찬나무방', and Settings page. In '감정놀이방', users engage with tasks to learn about different emotions. They are prompted to mimic expressions shown on the screen, and these expressions are evaluated for correctness through a webcam linked to the YOLO model. '행동놀이방' functions in a similar way, teaching the user appropriate recognition in the things rather than emotions. The Settings page allows users to customize features that affect '간식보상방'. On '칭찬나무방', users can view their progress in the form of apple-imaged points accrued through completed activities. This page also provides access to '노래보상방' and '간식보상방'. '노래보상방' allows users to spend their points on enjoying one of eight different content options. Similarly, '간식보상방' provides users with a choice of five different snacks that can be 'purchased' using their points.

### 3.5.2 Connect with back-end

In our front-end implementation using HTML, CSS, JS, PHP, and MySQL,


**Fig. 7.** UI example of the 칭찬나무방, 노래보상방, 간식보상방

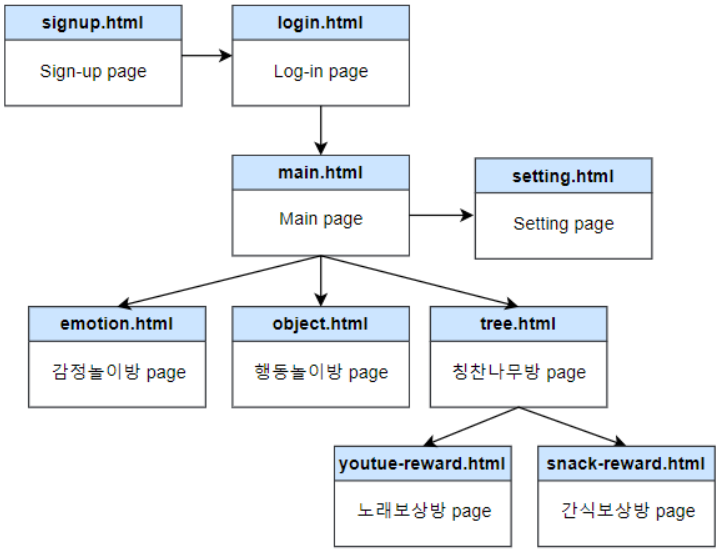


**Fig. 8.** Front-end architecture

we established connections between the front-end and back-end using the XMLHttpRequest object. The XMLHttpRequest object, a built-in feature in JavaScript, allows us to make HTTP requests from the browser. We chose to use the XMLHttpRequest object due to its compatibility across various browsers and its long-standing usage in web development. It provides us with flexibility and control over the request process, allowing us to set headers, handle response data, and manage request states. By using the XMLHttpRequest object, we ensured compatibility and maintained control over the request process in our front-end implementation.

### 3.6 Back-end detail

The API specification provided by the back-end is as follows.

```
var xhr = new XMLHttpRequest();
var url = 'setting.php';
var method = 'POST';

xhr.open(method, url, true);
xhr.setRequestHeader('Content-Type', 'application/x-www-form-urlencoded');

var requestBody = 'favsnack=' + encodeURIComponent(favoriteSnack) + '&userID=' + encodeURIComponent(userID)

xhr.onload = function () {
  if (xhr.status === 200) {
    var response = xhr.responseText;
    console.log(response);
    handleResponse(response); // Call a function to handle the response
  } else {
    console.error('Request failed. Status:', xhr.status);
  }
};

xhr.send(requestBody);
```

```
<?php
if ($_SERVER["REQUEST_METHOD"] === "POST") {
    if (isset($_POST["favsnack"], $_POST["userID"])) {

        $favsnack = $_POST["favsnack"];
        $userID = $_POST["userID"];

        // Establish a database connection
        $servername = "127.0.0.1";
        $username = "root";
        $password = "changethis";
        $dbname = "capstonedesign";
        $conn = new mysqli($servername, $username, $password, $dbname);

        // Check connection
        if ($conn->connect_error) {
            die("Connection failed: " . $conn->connect_error);
        }

        // Prepare the SQL update statement
        $sqlUpdate = 'UPDATE users SET favsnack = ? WHERE userid = ?';

        // Create a prepared statement
        $stmt = $conn->prepare($sqlUpdate);

        // Bind the parameters
        $stmt->bind_param('ss', $favsnack, $userID);
```

**Fig. 9.** setting.js and setting.php

**Table 1.** API specification.

| Index | Method | URL | Description |
|---|---|---|---|
| 1 | POST | signup/ | Sign up |
| 2 | POST | login/ | Log in |
| 3 | POST | emotion/ | Add sticker for the correct answer |
| 4 | POST | object/ | Add sticker for the correct answer |
| 5 | POST | word/ | Add sticker for the correct answer |
| 6 | POST | setting/ | Set child's customized character |
| 7 | GET | tree/ | Display the number of stickers |
| 8 | GET | youtube-reward// | Pay for the Youtube reward |
| 9 | GET | snack-reward/ | Pay for the snack reward |

## 4 Evaluation

As performance evaluation items for AI models, F1 Score, Precision-Recall, and mAP-0.5 were used. Precision can be calculated by using the ratio of the correct answers among the predictions of positive. And recall can be calculated by using the ratio of finding the real one out of the actual positives model have. And F1 score is a harmonized mean of Precision and Recall. mAP, mean average precision, means the ratio of the overlap of the ground truth and prediction. So, if mAP is bigger than 50%, consider it as a correct prediction.

For '감정놀이방' AI model, we did 100 epochs training with 16 batch size and 640 pixel image size. We got 93% F1 score(at 0.504 confidence), and 94% mAP-50 score. And for '행동놀이방' AI model, we did 150 epochs training with 64 batch size and 640 pixel image size. We got 56% F1 score, and 56% mAP-50 score.

# 5 Limitation And Discussions

## 5.1 AI

For the '감정놀이방' Dataset,

1. In the case of emotions without a clear characteristic difference, it was not well distinguished, so that emotion class was deleted.
2. Since most of the raw data are Western images, the performance may be slightly lower than expected results because the ratio of Asian images has been trained significantly less.
3. The users of our project are mainly autistic patients from childhood to adolescence, and raw data mainly focuses on adult images, so it recognizes children's facial emotions, and the performance is bound to be lower than expected.

For the '행동놀이방' Dataset,

1. The accuracy didn't come out as high as we thought because there is a limit to getting a high-definition quality dataset.
2. For the above reasons, it is often not recognized according to the angle at which the object appears.

## 5.2 Web

Not all web browsers offer the same functionality, and some browsers may not support the latest web technology. This can change the user experience, and web applications should provide compatibility to as many users as possible. Second, real-time video streaming can be vulnerable to network problems such as bandwidth, latency, and data packet loss. This can reduce video quality, which can negatively affect the user experience. Finally, these kinds of applications impose important considerations for security and privacy. This information should be properly protected because users share personal data through the webcam. This requires security protocols, encryption technology, and strict policies regarding data storage and processing.

# 6 Related works

## 6.1 Mobile application

'다은' is an application that helps authentic children learn how to imitate emotions. It consists of the following 3 steps. Authentic children follow the emotion or movement in the picture and they take a photo by tapping the camera button on the right side. Finally they upload the taken photo from their photo album. Key features include: emotion imitation and movement imitation. However, taking a photo and uploading it from their photo album is difficult for authentic children and it can degrade their motivation for learning.

## 7 Conclusion

We planned and developed programs to help children with ASD develop the skills needed for social adaptation and interaction by leveraging AI's skills and effective learning skills to help alleviate some of the challenges facing them and their families. To develop the program, we tried to solve the problem using AI technologies such as Computer Vision, YOLOv8, and various types of Dataset, as well as tools such as HTML, CSS, and JS in the front-end part and Python, Flask, Ngrok, and GCP in the back-end part. The first method was the web application method, but it was completed as a package program for technical reasons of webcam connection. However, we believe that we have overcome the limitations of existing education methods by providing a more accessible and effective education platform, and we hope we improved the learning environment and quality by providing more attractive and effective learning experiences for ASD children.

## 8 References

1. https://github.com/ultralytics/ultralytics
2. https://blog.roboflow.com/whats-new-in-yolov8/
3. https://app.roboflow.com/nayoung-79nh6/emotion2-tc0yj/deploy/2
4. https://metahuman.unrealengine.com/mhc
5. https://clovadubbing.naver.com/
6. https://m.riss.kr/search/detail/DetailView.do?p$_m at_t ype = be54d9b8bc7cdb09control_n o =$ 6c15c606a71103eaffe0bdc3ef48d419