Intervention Device for Autistic Children's Tantrum Based on Artificial Intelligence

Nyayu Latifah Husni[1,*], Selfi Nurhaliza[2] Dewi Permata Sari[3], Masayu Anisah[4], Abdurrahman[5], Ade Silvia Handayani[6] , Charina Mutiara Chairunnisa[7], Syarif Husin[8]

[1-8] Politeknik Negeri Sriwijaya, Palembang 30139, Indonesia

nyayu_latifah@polsri.ac.id

Abstract. Children with autism often face behavioral challenges, including tantrum behaviors caused by difficulties in communicating their needs or desires. Tantrums are excessive emotional reactions resulting from the inability to control and communicate feelings. Therefore, an Artificial Intelligence (AI)-based system is needed to detect and intervene in tantrum behaviors. The libraries used in developing this system are OpenCV and MediaPipe. OpenCV captures and processes images from the webcam, while MediaPipe analyzes movements using Python-based algorithms. MediaPipe tracks body landmarks in real-time (numbers 15 and 16 for hands; numbers 27 and 28 for feet). If a tantrum is detected, intervention is performed by playing music to soothe the child. It is categorized into several levels (1, 2, and 3). The mini PC will signal the speaker to play music corresponding to the detected tantrum level.

Keywords: Autism, Tantrum Behavior, MediaPipe, Mini PC, Speaker, Music, Artificial Intelligence

1 Introduction

Autism, known as Autism Spectrum Disorder (ASD), is a pervasive developmental disorder characterized by impairments and delays in communication, language, social interaction, interests, and behavior [1]. The number of autism cases in Canada and Japan has increased by 40% since 1980; in California, in 2002, it was concluded that there were 9 cases of autism per day, and in [2] the prevelance estimates indicate that in 1in 54 children in the United States estimated has ASD. The prevalence of ASD in Southeast Asia has been determined to be six cases per 1000 individuals, with a greater prevalence among males in comparison to females [3]. If the birth rate in Indonesia is six million per year, then the number of autistic people in Indonesia increases by 0.15% or 6,900 children per year. The prevalence of boys is three to four times greater than girls [4].

Autistic children often experience behavioral difficulties compared to other children in general, facing behavioral challenges that can affect their daily lives and those around them. One of the main challenges autistic children face is the emergence of tantrum behavior, which is a form of inability to express their needs or wants in a way that others can understand. Tantrums in autistic children often cause stress for the child and those around them, including parents or caregivers. Therefore, a system is needed to detect and intervene in tantrum behavior as an effort for parents or caregivers to manage the situation [5].

With previous research related to the system for detecting early symptoms of tantrums in autistic children through facial expression analysis using the Deep Face method, developed the research into a system that can detect tantrum movements and provide interventions during tantrums using songs in autistic children to help parents or caregivers. Based on this, took out the research title "Artificial Intelligence-Based for Intervention Device Autistic Children's Tantrum".

Artificial intelligence can be utilized to detect and intervene tantrums in autistic children. One of the libraries that can be used to develop this system is MediaPipe. MediaPipe automatically detects and tracks landmarks on the body in real-time, providing essential data that AI algorithms can process for motion analysis. In this case, landmark numbers 15 and 16 are used to track the hands, while numbers 27 and 28 are used to track the feet. By using these landmarks, the system can accurately monitor body movements. With real-time data from MediaPipe, the AI system can detect tantrums based on rapid changes in movement patterns. If a tantrum is detected, an intervention will be made using songs to calm the autistic child.

N. L. Husni et al. (eds.), *Proceedings of the 8th FIRST 2024 International Conference on Global Innovations (FIRST-ESCSI 2024)*, Advances in Engineering Research 261,
https://doi.org/10.2991/978-94-6463-678-9_38

2 State of The Art

Artificial intelligence is a field of computer science that focuses on creating computer systems that can perform tasks that usually require human intelligence. AI is used to solve complex problems quickly and efficiently, especially those that require extensive data analysis. One subset of AI is Machine Learning. Machine Learning focuses on using data and algorithms to mimic how humans learn and gradually improve accuracy, characterized by training on a dataset and then prediction to estimate the output of a dataset based on data learned during training [6].

Human Pose Estimation can be defined as the process of localizing human joints or predefined landmarks. Figure 1 is a landmark model (Blaze GHUM 3D) as a marker showing the main joints and locations of the human body. The input image is fed to the Mediapipe BlazePose library to detect key points on the user's body. The output is a list of coordinates on the X, Y, and Z axes for 33 major key points of the human body. The primary purpose of adopting pose estimation is to track those key points in a video or photo [7]. A webcam is used to capture the video, then the video is displayed in OpenCV. Then, the video is processed for each key point by pose estimation using the Mediapipe library for detection, with the output being a real-time body skeleton. The code suite is written in Visual Studio Code using Python.

MediaPipe is a working platform by Google that develops machine learning solutions on Android, the web, and desktops. MediaPipe is designed to track a person's movements, so each body part will display a line of connecting dots. Featured features include object detection, face recognition, pose estimation, and real-time hand detection.

OpenCV is an open-source library designed specifically for displaying images or videos. Computer vision is a branch of image processing that allows computers to see like humans. With this ability, computers can make decisions, perform actions, and recognize particular objects. Meanwhile, Image Processing is a method of processing input data from two-dimensional images into output data or information that has undergone processing. Image Processing processes capture images and then convert the data so humans can read it, which is done by a computer [8].

The programming language that can be used is Python. Python is a programming language that uses an interpreter to run its program code. Python can be applied in several applications, such as image processing, web applications, software development, audio or video-based applications, et cetera.

3 Methods

The tantrum detection and intervention system has three important parts: input, process, and output. The input is in the form of images captured in real-time using a webcam, then directly processed by OpenCV. Meanwhile, MediaPipe is the brain that analyzes the detected movements using Python-based algorithms. The result of motion detection from the captured object, the data will be sent to the Mini PC for processing. Based on the results of the analysis, the Mini PC will signal the speaker to play a song that matches the number of tantrums detected. The block diagram of the tantrum detection and intervention system is shown in the figure below.

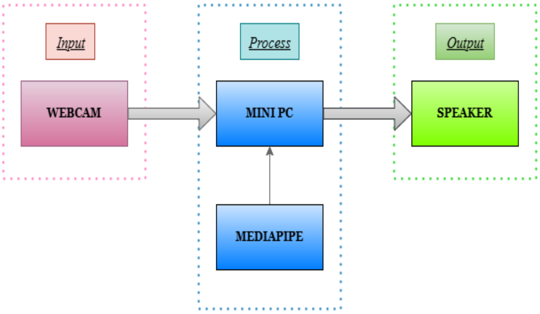


Figure 1 Block Diagram of Tantrum Detection and Intervention System

The figure below displays the hardware of the tantrum detection and intervention system consisting of a webcam, mini PC, speaker, and laptop.

Figure 2 Hardware

Song levels are grouped into three categories: level 1 songs, level 2 songs, and level 3 songs. A level 1 song will come out if three tantrums are detected, which means that every time the system detects a tantrum three times, level 1 song will be played. Level 2 song will be played if six tantrums are detected, meaning that every time the system detects a tantrum up to six times, a level 2 song will be played. The level 3 song will come out if nine tantrums are detected, meaning that every time the tantrum detection reaches nine times, the level 3 song will be played. The system's workings can be seen in the flowchart below.

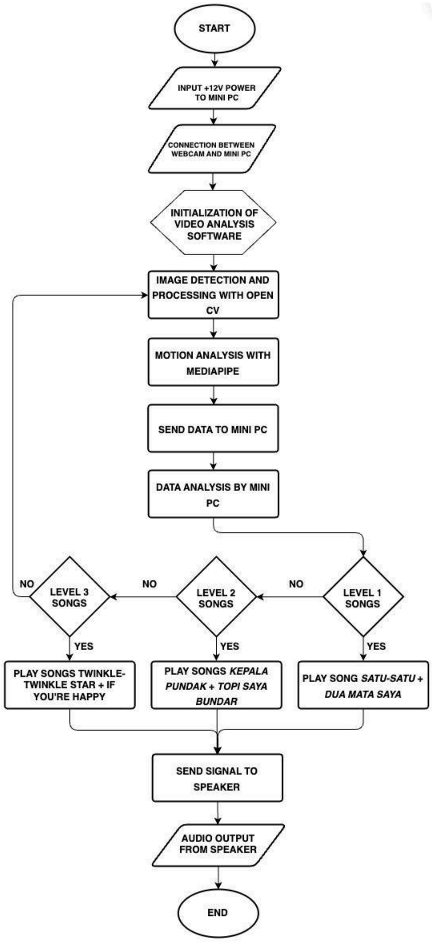


Figure 3 Tantrum Detection and Intervention System Flowchart

The distance between the detection and intervention system for autistic children is 1-1.5 meters. If the system detects a tantrum on the object, a song is played through the speakers to overcome the tantrum, as shown in Figure 5.

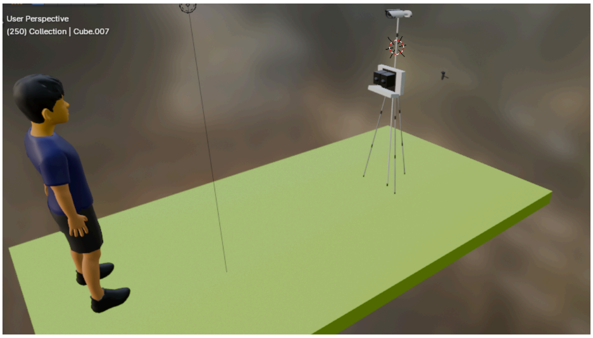


Figure 4 Distance Implementation Design on The System

4 Result

The system detects movement with an effective distance of 126 cm, with the body positioned straight ahead. This effective distance shows the system's ability to respond to movement with good accuracy and responsiveness, as shown in Figure 6 below.

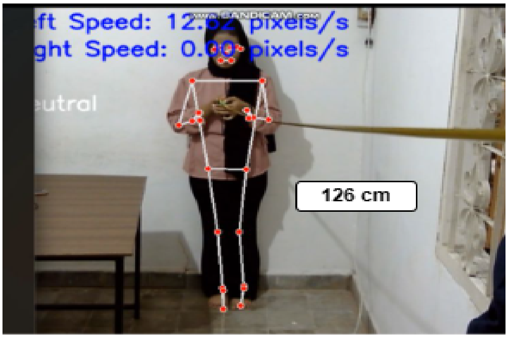
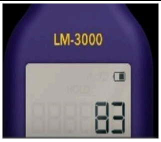
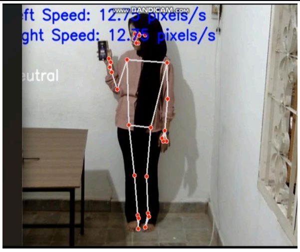
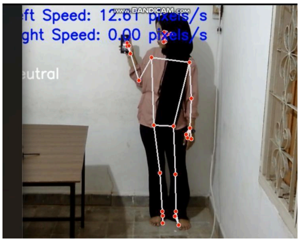
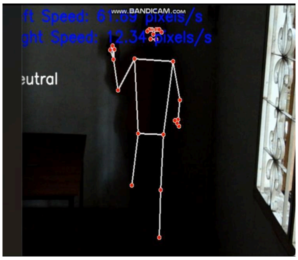


Figure 5 Motion Detection Range

Table 1 below tested the light intensity variations in various conditions using the Light Meter application with LUX units. The system was tested to detect objects for each condition that produces a light intensity value. The results showed that the system is able to perform object detection effectively in conditions with a certain light intensity, with varying success rates depending on the lighting level. Based on several given conditions, both high and low light intensity greatly affect the performance of the system.

Table 1 Results of Light Intensity Testing

No.	Light Intensity (LUX)	Detection Results	Description

1	 Condition: Lights on and window open		Detected
2	 Condition: Lights on and window close		Detected
3	 Condition: Lights off and window open		Not Detected
4	 Condition: Lights off and window close		Not Detected

Testing of movement conditions was carried out with various variations, such as hands forward, one hand up, two hands up, sitting, and sleeping. Tests were conducted with fast movements to test the system's responsiveness and accuracy to various types of movements. The results of these tests provide a better understanding of the system's ability to detect and respond to movements of varying complexity and speed. Table 2 shows the results of testing the movement conditions performed by the object.

Table 2 Motion Condition Test Results

No.	Motion	Description
1	 Condition : Hand Forward	Detected Left Speed: 101.28 pixels/s Right Speed: 105.51 pixels/s Left Ankle Speed: 0.00 pixels/s Right Ankle Speed: 7.07 pixels/s
2	 Condition : One hand up	Detected Left Speed: 8.75 pixels/s Right Speed: 152.73 pixels/s Left Ankle Speed: 0.00 pixels/s Right Ankle Speed: 4.37 pixels/s
3	 Condition : Two hands up	Detected Left Speed: 74.30 pixels/s Right Speed: 134.56 pixels/s Left Ankle Speed: 11.08 pixels/s Right Ankle Speed: 19.81 pixels/s
4	 Condition : Sitting	Detected Left Ankle Speed : 58.24 pixels/s Right Ankle Speed: 117.07 pixels/s Left Ankle Speed: 26.89 pixels/s Right Ankle Speed: 11.79 pixels/s

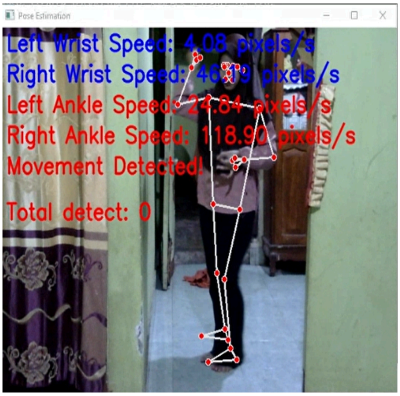
5	 Condition : Lying Down	Detected Left Ankle Speed : 58.76 pixels/s Right Ankle Speed: 165.00 pixels/s Left Ankle Speed: 375.21 pixels/s Right Ankle Speed: 335.00 pixels/s

Testing motion detection at several-degree angles helps measure the reliability and accuracy of the motion detection system under various viewing angle conditions. To test the effective angle of motion detection, a wide variety of angles between 0-90° to the left and 0-90° to the right were tested. The angles were measured using an arc, with the body pointed at the predetermined angle.

Table 3 Motion Detection Testing at Various Degree Angles

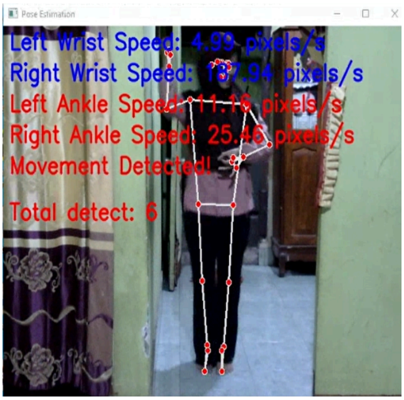
Angle Position	Results	Detection Description
90° to the rigth		Maximal Left Wrist Speed: 25.02 pixels/s Right Wrist Speed: 138.58 pixels/s Left Ankle Speed: 156.13 pixels/s Right Ankle Speed: 58.50 pixels/s
60° to the rigth		Maximal Left Wrist Speed: 70.14 pixels/s Right Wrist Speed: 267.46 pixels/s Left Ankle Speed: 47.22 pixels/s Right Ankle Speed: 15.81 pixels/s

30° to the rigth



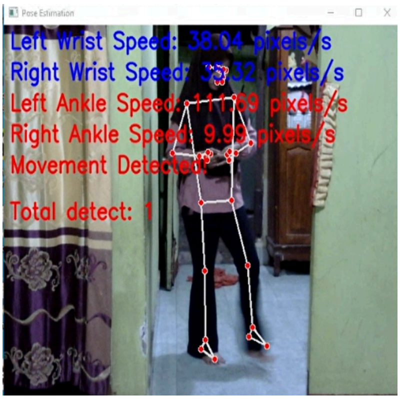
Maximal
Left Wrist Speed: 4.08 pixels/s
Right Wrist Speed: 46.19 pixels/s
Left Ankle Speed: 24.84 pixels/s
Right Ankle Speed: 118.90 pixels/s

0° to the front



Maximal
Left Wrist Speed: 4.99 pixels/s
Right Wrist Speed: 187.94 pixels/s
Left Ankle Speed: 11.16 pixels/s
Right Ankle Speed: 25.46 pixels/s

30° to the left



Maximal
Left Wrist Speed: 38.04 pixels/s
Right Wrist Speed: 35.32 pixels/s
Left Ankle Speed: 111.69 pixels/s
Right Ankle Speed: 9.99 pixels/s

60° to the left		Maximal Left Wrist Speed: 80.03 pixels/s Right Wrist Speed: 15.47 pixels/s Left Ankle Speed: 172.92 pixels/s Right Ankle Speed: 39.26 pixels/s
90° to the left		Maximal Left Wrist Speed: 717.53 pixels/s Right Wrist Speed: 92.11 pixels/s Left Ankle Speed: 5.11 pixels/s Right Ankle Speed: 221.36 pixels/s

The system is able to maximally detect movement at all angles from 0-90° both to the right and to the left. If the detection is not optimal, it means that the system does not detect the landmark points on the hands and feet well. The results of the motion detection test are assessed based on the quality of detection of landmark points on the body. If the maximum detection caption is achieved, all landmark points, such as hands and feet, are completely and accurately detected by the system. In other words, the detection system captures all relevant body parts without missing any. Conversely, if the detection caption is not maximally achieved, it indicates that the system did not detect some landmark points on the body. This means that some body parts were missing or not identified.

The intervention system using songs for autistic children is a method that aims to calm and reduce anxiety in autistic children and can be done by dividing songs into levels. In the first level, the songs used are "*Satu-Satu*" and "*Dua Mata Saya*". Furthermore, in the second level, the songs used are "*Kepala Pundak Lutut Kaki*" and "*Topi Saya Bundar*". In the third level, the songs used are "Twinkle-twinkle Little Star" and "If You're Happy and You Know It".

Table 4 Results of Testing the Effectiveness of Songs

Song	Expression	Singing	Movement	Detection Result

Level 1	Enthusiastic	Yes	Yes	
Level 2	Enthusiastic	Yes	Yes	
Level 3	Flat Affect	Yes	Yes	

The autistic child's response to some of the songs given by the author, there were signs such as:

1. Expression

   When hearing level 1 and 2 songs, the autistic child shows enthusiastic facial expressions such as smiling or laughing. However, when hearing level 3 songs, the child shows a flat or ordinary facial expression, known as "flat affect". A "flat affect" expression refers to a face that shows no emotion when someone feels something. So, the autistic child may feel something positive but not express it through a smiling or laughing facial expression.

2. Singing

   The autistic child sings along by babbling when they hear Level 1, 2, and 3 songs. The child tries to imitate the melody or lyrics of the song being played.

3. Moving

   Autistic child move their hands, feet, and head when they hear songs in levels 1, 2, and 3. Autistic child can show different responses to different types of songs due to several factors. In the first level, the songs used are "*Satu-Satu*" and "*Dua Mata Saya*". Furthermore, in the second level, the songs used are "*Kepala Pundak Lutut Kaki*" and "*Topi Saya Bundar*". These songs are more familiar and easy to understand, thus making autistic children express themselves enthusiastically. In addition, the songs have faster rhythms and more interactive lyrics, which can encourage children to move and express themselves more actively. In contrast, English songs may not be as familiar and may have a slower tempo or softer melody, which tends to calm the child down and produce a quieter, flatter response.

5 Conclusion

In conclusion, the results of the research that the system effectively detects movement at a distance of 126 cm when the body is positioned directly in front. It was observed that both high and low light intensities significantly influence the system's performance. Additionally, the device demonstrates optimal movement detection across all angles ranging from 0 to 90°, effectively recognizing movements to the right and left. It can identify various movement conditions, including hands extended forward, one hand raised, both hands raised, as well as sitting and lying down positions. Furthermore, it was noted that songs with clear lyrics, fast tempos, and interactive elements tend to elicit more enthusiastic reactions from autistic children, while slower English songs with softer melodies often result in more subdued expressions.

Reference

[1] J. L. Matson and P. Sturmey, “Autism and Pervasive Developmental Disorders.”

[2] R. R. Redfield *et al.*, “Morbidity and Mortality Weekly Report Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years-Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2016 Centers for Disease Control and Prevention MMWR Editorial and Production Staff (Serials) MMWR Editorial Board,” 2020.

[3] M. Shrestha *et al.*, “Prevalence of autism spectrum disorder among children in Southeast Asia from 2002 to 2022: An updated systematic review and meta-analysis,” *Health Sci Rep*, vol. 7, no. 4, Apr. 2024, doi: 10.1002/hsr2.2005.

[4] S. R. Arumugam, R. Balakrishna, R. Khilar, O. Manoj, and C. S. Shylaja, “Prediction of Autism Spectrum Disorder in Children using Face Recognition,” in *Proceedings - 2nd International Conference on Smart Electronics and Communication, ICOSEC 2021*, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 1246–1250. doi: 10.1109/ICOSEC51865.2021.9591679.

[5] G. Nie *et al.*, “An Immersive Computer-Mediated Caregiver-Child Interaction System for Young Children with Autism Spectrum Disorder,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 29, pp. 884–893, 2021, doi: 10.1109/TNSRE.2021.3077480.

[6] B. Kamala, K. S. Mahanaga Pooja, S. Varsha, and K. Sivapriya, “ML Based Approach to Detect Autism Spectrum Disorder (ASD),” in *Proceedings of the 2021 4th International Conference on Computing and Communications Technologies, ICCCT 2021*, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 313–318. doi: 10.1109/ICCCT53315.2021.9711826.

[7] S. Negi, M. Garg, H. Maindola, V. Kansal, U. Jain, and S. Bhatla, “Real-Time Human Pose Estimation: A MediaPipe and Python Approach for 3D Detection and Classification,” in *Proceedings - International Conference on Technological Advancements in Computational Sciences, ICTACS 2023*, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 128–133. doi: 10.1109/ICTACS59847.2023.10390506.

[8] Y. Zhang and X. Zheng, “Development of Image Processing Based on Deep Learning Algorithm,” in *2022 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers, IPEC 2022*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 1226–1228. doi: 10.1109/IPEC54454.2022.9777479.