Artificial Vision Algorithm for Behavior Recognition in Children with ADHD in a Smart Home Environment

Jonnathan Berrezueta-Guzman[1], Stephan Krusche[1], Luis Serpa-Andrade[2], and María-Luisa Martín-Ruiz[3]

[1] Technische Universität München, 80333 Munich, Germany
[2] Universidad Politécnica Salesiana, 010105 Cuenca, Ecuador
[3]Universidad Politécnica de Madrid, 28031 Madrid, Spain
s.berrezueta@tum.de, krusche@in.tum.de, lserpa@ups.edu.ec, marialuisa.martinr@upm.es

Abstract. Artificial vision has made a great advance in the recognition of visual patterns that are not perceptible by humans or that are biased in their interpretation. Among its applications, artificial vision or computer vision has served in the support of people with some kind of disability. In this work, an image classification algorithm is developed to complement a pervasive therapy support system for children with Attention Deficit Hyperactivity Disorder (ADHD) during the development of their homework. For this purpose, a camera is adapted within a smart environment made up of Smart objects and a robotic assistant. In the system, a convolutional neural network (CNN) is implemented for the classification of the child's status (doing or not doing his/her homework). An experiment of this implementation is carried out in which the results of the environment without the camera are compared with the results obtained by using the camera and the implemented CNN. The latter results are also compared with the information collected through observation by the therapist during the session. The results show that what the camera identifies as the child not doing homework matches what the smart objects identify as distractions and pauses at 82.70% and what the therapist identifies as distractions and pauses at 98.21%. This approach will help the smart home environment have new and more accurate data to process and make better decisions, just like a therapist would do.

Keywords: Artificial Vision, Computer Vision, Convolutional Neural Networks, Robotics, Smart Home, ADHD.

1 Introduction

Identification of the environment has allowed humans to develop the skills to manipulate it to their benefit. The sense of sight has been classified as the most important because its application is primordial for humans when making decisions [1]. By implementing this sense in a machine, its information processing potential could be exploited to determine patterns and classify information that the human eye could not, in addition to providing more utilities so that the machine can also make fast and accurate decisions [2].

In recent years, it has been evident that machine vision has meant a significant advance in image classification research with a speed and accuracy that are immeasurably greater than that of the human eye [3]. Today, it is easy to access various types of classification algorithms and manipulate them to orient their functionality to meet a specific objective [4, 5].

This article details the process of adapting an image classification algorithm in an intelligent environment that could help in the occupational therapy processes of children with ADHD at home. In section 2, the related work is described, as well as the results obtained and the possible improvements that this new implementation could bring. Section 3 details the applied methodology that includes the hardware design for the implementation of the camera in the robot structure, the design and development of the artificial vision algorithm in the main board of the robot, the tests performed and the experiment where the results obtained by the camera are compared with the results obtained by the intelligent objects and the therapist. In section 4, the results obtained in the experiment are discussed, and in the conclusions section, the authors detail what has been the contribution of this development and what will be future work.

2 Related Work

ADHD is a disorder that affects millions of children around the world [6]. Its causes are varied and its diagnosis has been controversial in recent years due to its comorbidity and association with other neurodevelopmental disorders [7]. Several previous works have applied artificial vision in the field of support for people with disabilities [8], other approaches have been developed in collaboration with robotics and other emerging technologies such as virtual reality and augmented reality [9-11].

In previous publications [12] we have developed an accompaniment and support environment is developed for children with ADHD as a therapy process along with homework sessions. This environment consists of objects connected to the Internet that help identify behavioral patterns by measuring physical parameters such as movement and distance of objects from the child. These parameters indicate whether the child plays with the chair or leaves the workstation. In addition, it has a robotic assistant as feedback that acts as a therapist: guiding and correcting the child's behavior through motivational phrases and expressions, providing a human-machine interaction (see figure 1).

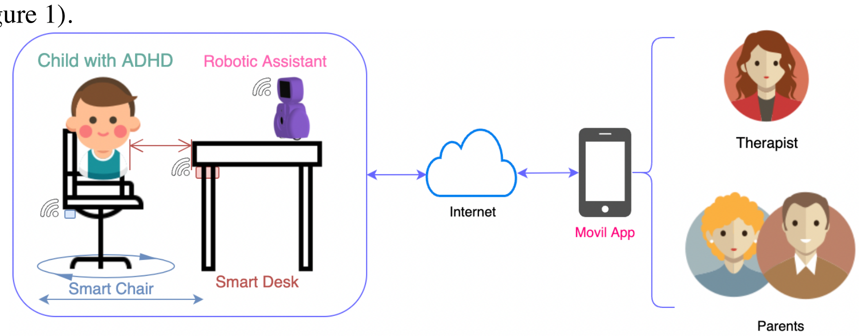

Fig 1. The intelligent environment where computer vision will be implemented to determine the behavioral parameters of children with ADHD [13].

We have evaluated the entire environment with the robotic assistant [14] and compared the results that the children had during eight weeks with this environment and without it. A good evolution was evidenced during the sessions in which the children worked in the smart-home environment. In addition, several parameters such as distractions, pauses, attendances, and hyperactivity were detected at a high level of coincidence compared to what was also detected by the therapist.

However, the same study identifies that there are patterns that the smart objects in this environment cannot identify but the therapist can, such as the child's mood change and small-scale impulsivity (playing with objects on the desk and making sounds with objects). In addition, even though there is no movement of the chair and the child's presence at the workstation does not change, it is not possible to know if he/she is actually doing his/her homework or not engaged with it. However, such parameters in the experiment were perceived and documented by the therapist. In contrast, neither the Smart objects nor the robot could determine these.

The research question posed by this paper is whether a computer vision algorithm can identify and interpret the behavioral patterns of a child with ADHD. In this first approach, we only try to recognize the parameter of whether or not the child is doing homework on his/her desk. For this purpose, we propose to use computer vision to classify images of the child's desk and the items on the desk that interact with him/her.

3 Methodology

When implementing artificial vision in the smart home environment, it is necessary to introduce a small camera into it; however, it is important to avoid the child feeling spied on or watched. Intelligent objects and the robotic assistant do not cause this sensation in the child in [14] and the implementation of a camera could have the opposite effect. However, by evaluating the structure and composition of the hardware of the robotic assistant, it is evident that it is possible to adapt a small camera to it.

3.1 Hardware Customizability

In its composition, the robot has a Raspberry pi 4 mini-computer board that allows connecting a Raspberry Pi Camera Module 2 NoIR which has a Sony IMX219 8 megapixel sensor. This camera does not employ an infrared filter (NoIR), which makes photos taken in daylight or night light look better at any level of light intensity [15].
For this purpose, the camera is adapted to the front part of the robotic assistant. As shown in Figure 2, this camera goes unnoticed by the child. Adding this camera to the robot hardware does not represent the modification of the original robot design, as there is a slot in the front of the robot to attach some kind of sensor or additional component (in this case, the camera). The energization requirement is not affected because the consumption of the camera energy is neglectable even with a significantly low image capture period.

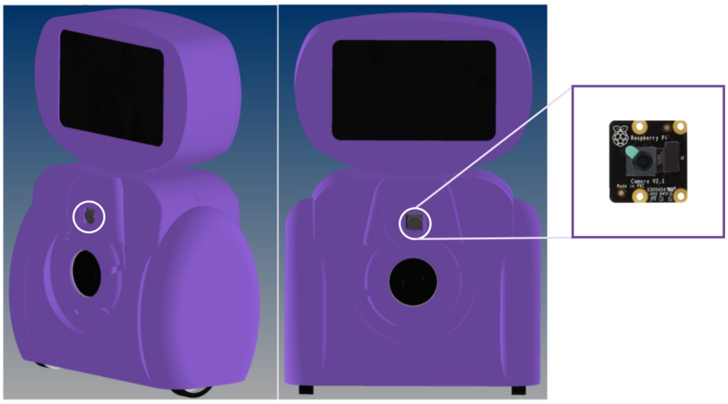

Fig 2. Design of the robotic assistant with its adapted camera on the front side. This design was inspired by [16] which performed well with school children in the process of teaching traumatic accident prevention.

3.2 Classification Algorithm

To recognize parameters that determine whether the child does homework or not, employing a camera embedded in the robot structure, it is proposed to deploy a computer vision algorithm with a machine learning approach using Convolutional Neural Networks (CNN). This algorithm will be implemented in parallel to the functionalities already established in the robot software. Therefore, Figure 3 shows the hardware and software components developed in the Raspberry Pi inside the robot: intelligence, peripheral management, motor control, internal storage, memory access, and its communication protocol for information transfer and command reception [17].

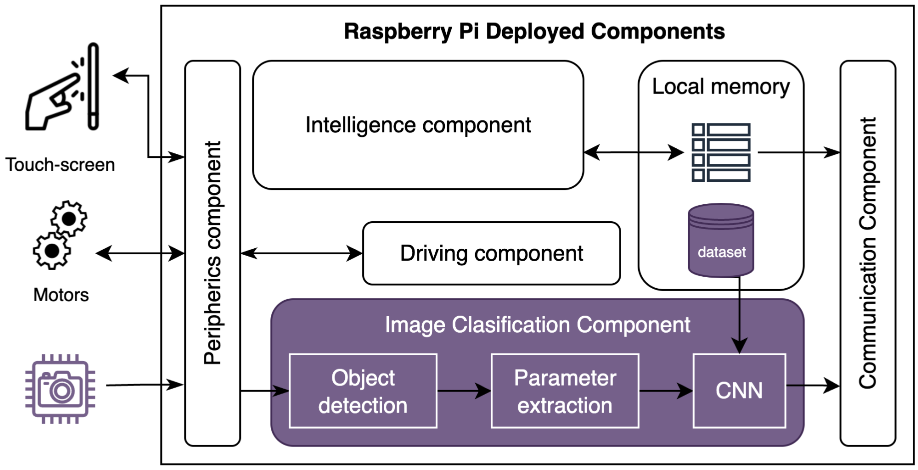

Fig 3. Implementation of the image classification algorithm inside the raspberry board.

Figure 3 shows the new image classification component which starts capturing images with the camera connected to the peripheral controller. Following is explained the parts of this new component.

Dataset. First, a set of data (images of children doing homework) is collected. These images are related to typical writing elements, such as images of pencils/pens of all types and sheets of paper on a surface. In addition, images are provided of the pencil being held in the hand and this, in turn, is in contact at different angles with the paper. These images were taken in therapy sessions with ADHD children in an intelligent environment while they were doing their homework. The images were taken every 10 seconds, so after 8 weeks we have a dataset of 5262 images in which pens and pencils are observed held in the hand (left or right), not held, or none of these pencils are observed. Similarly, there are also images in which the absence of the child from the work environment is evident.

This data set collects several working conditions, such as different levels of occlusion, lighting, and blurring. In addition, the images contain other objects on the desk and various patterns on the children's clothes that could impact the performance of the algorithm. All these conditions were established to provide a diverse set of images to train the algorithm and cover all possible negative patterns.

All images were taken without revealing the identity of the child, as the focus area is from the neck to the core of the child. Table 1 shows how these images would be classified for interpretation.

Table 1. Interpretation of pattern recognition.

Object	Pattern recognition	Interpretation (<30 s)	Interpretation (>30 s)
Pencil	On the hand & on the paper	Doing the homework	Doing the homework
	In the hand & away from the paper.	Pause the homework	Not doing homework
	Not in the hand & on the paper.	Pause the homework	Not doing homework
Hands	No hands on the desk	Pause the homework	Not doing homework
	Holding an unknown object	Pause the homework	Not doing homework
	Next to the paper	Pause the homework	Ask for feedback.

The time in seconds can be modified according to the estimation made by the therapist with a previous evaluation with the children. All these interpretations will be compared with the results that the therapist receives in the experiment and also with the information collected from the smart objects.

Implementation of the algorithm.

i) The Raspberry Pi implements a Python software to continuously capture images to be classified using the camera. Each capture is made at 10-second intervals;

ii) The sample is divided into training (3684 images = 70%) and testing data sets (1578 images = 30%).

iii) Small errors in image comparison could be caused by occlusions, blur, lightning, and movements that affect the quality of the result. To avoid these bad results, data augmentation is applied to improve the quality of the classification;

iv) Finally, the convolutional neural network is run to classify the data set as child *does homework* or *does not*. CNN uses feature extraction of features related to the activity of doing homework on the desk: pencil in hand at a certain angle with interaction with the paper [18].

The CNN used has a simple architecture with three convolutional layers and three fully connected layers as shown in Figure 4.

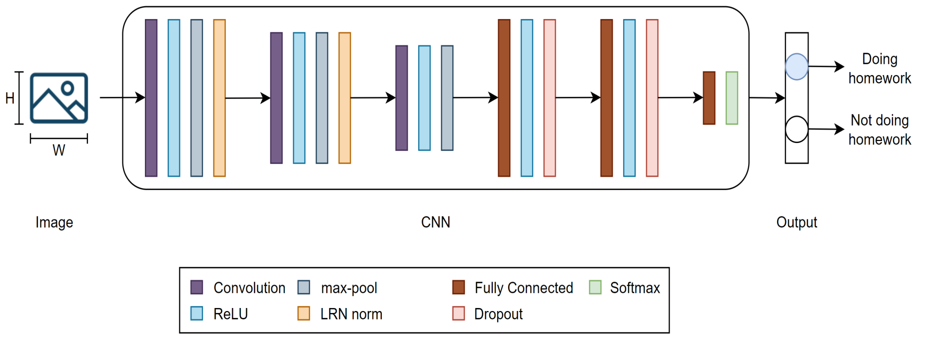

Fig 4. Convolutional neural network architecture with three convolutional layers and three fully connected layers.

A data augmentation layer was implemented in the input of the network. The data augmentation included random reflection, rotation, and zooming in each forward and backward training image for reducing the over-fitting.

To obtain behavior recognition if the child is doing the homework or not, it was necessary to crop to 700 x 700 pixels around the centre of the image, which is used to feed the network.

Convolutional Layer One. This layer contains 96 filters of 3 x 7 x 7. These filters are applied to the acquired image through the camera inside the robot. Following, four strides and zero-padding are specified. To reduce the output size, this is followed by a linear rectified unit (ReLU) and max-pooling. At the end of this layer, a local response normalisation (LRN) is applied to improve contrast enhancement to use the maximum values of pixels as excitation for the next layers [19].

LRN supports the generalization of CNNs and helps to introduce lateral inhibition between the different filters for the given convolution.

Convolutional Layer Two. Following the output of the first layer, the second convolutional layer is applied with 256 filters and a size of 96 x 5 x 5, followed by one stride and padding equal to two. Just like the first convolutional layer, ReLU, max-pool, and LRN are applied to reduce the size of the output and to improve the contrast enhancement for the next layers.

Convolutional layer three. This layer is applied to the output of the second layer with 384 filters of size 256 x 3 x 3 followed by 1 stride and 1 padding. In this layer there is

no LRN because at this point the output is more clear and it is more evident the classification of the image.
Fully Connected Layers. After the convolutional neural network, three fully connected consecutive layers are applied. The first receives the output from the third layer of the convolutional neural network, and the second receives the output of this first fully connected layer. Both fully connected layers have 512 neurons and are applied ReLU and dropout layers. The two dropout layers are entered with 0.5 of ratio to set to 0 the output value of the neuron.

The third fully connected layer contains 8 neurons and is connected to the output of the second fully connected layer. This layer maps the final indication of whether or not the child *does* or *does not do homework*. The dropout layer of the last fully connected layer is fed to a soft-max layer, which returns the calculated probability for each option.

This function is also known as multinomial logistic regression and is used to calculate the loss term and class probabilities during a classification. Furthermore, it is used to optimize training and minimise cross-entropy, which would be seen as one predicted for the real class and zero predicted for everything else [20]. The decent stochastic gradient optimizer is used for the dataset to train itself with a batch size of 50 images.

Testing. The algorithm returns when the subject (child) is doing the homework or not with a calculated probability. The result of this classification has a performance of 94,32%. The image (a), (b), and (c) in Figure 5 show a high probability of the child doing his homework because the captured image shows a hand holding the pencil over the paper, which can be interpreted as the child is doing the homework. On the other hand, image (d) in Figure 5 shows a low probability of the child doing the homework because the captured image shows a pencil over the paper in a certain angle and it is not held by a hand.

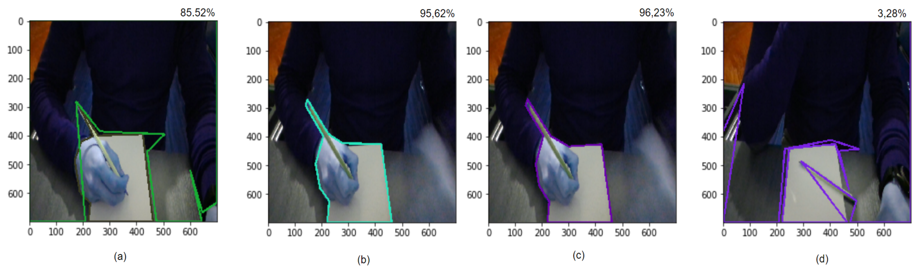

Fig 5. Determination of the probability that the subject (child) *is doing homework*.

3.3 Experiment

The hypothesis of this study is to be tested using a test and a comparison of results in an experiment with children with suspected ADHD.

- In the first stage, it is analysed whether the parameters detected by the Smart objects are related to the results of the child *not doing the homework* by the camera in the robot.

- In a second stage, a comparison is made as to whether what the robot interprets as "child not doing homework" coincides with what the therapist interprets as distractions and pauses.

For this experiment, the Conners [21] test is applied to 4 six-year-old children and a possible suspicion of ADHD is revealed. These children use the components of the Smart environment and do their homework in the company of the therapist (just like the experiment in [14]). The environment collects data on the child's behavior as well as the therapist during one hour sessions for 10 days.

4 Discussion of the results

The results of the experiment with the four children in the sample showed a direct relationship between what intelligent objects interpret as a distraction and pause with what the camera interprets if the child is not doing his/her homework. It can be argued that each child is different in the way they behave; in fact, if the objects detect distraction and the camera does not detect a state of not doing the homework, it may be because the child is still concentrating on the task, but their body is making involuntary movements in the chair, but in the end the concentration is still there.

However, it is not the same the other way around; if the state of not doing the homework is detected by the camera but the smart objects do not detect distraction or pause, it is definitely related to loss of concentration and even assumed that the child is wandering if this state does not change in a considerable time. This result is positive for this implementation because it shows the need for this sensor to measure this parameter and avoid biases.

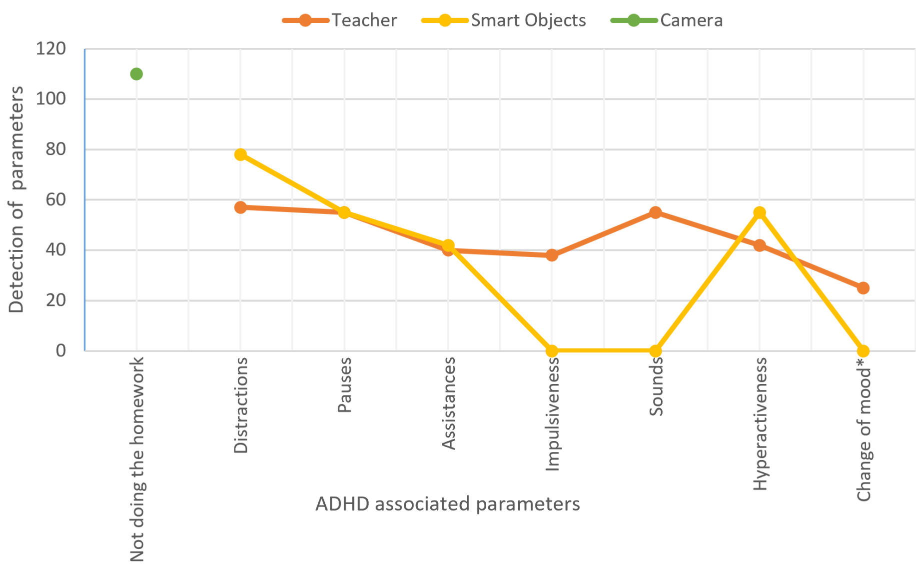

Fig 6. Results of the experiment. Comparison between the number of states not doing the homework with distractions and pauses detected by the therapista and the smart objects.

Figure 8 shows that the 110 times that the camera detected the child *not doing the homework* are directly with the 55 *distractions* plus the 57 *pauses* (total of 112 events) that the therapist detected (98,21 % of coincidence) and with the 78 distractions and 55 pauses (total of 133 events) that the smat objects detected (82,70 % of coincidence). The p-value calculated using a 95% confidence level in the two stages of the experiment is 0.027 and 0.0259 respectively. This means a high statistical significance in the parameters observed by the intelligent objects with the algorithm and the parameters observed by the therapist with the algorithm respectively.

Concerning the same detection, it is evident that the results coincide very strongly with what the therapist interprets. In fact, it is evident that the algorithm detects the event a little earlier than what the therapist detects. The therapist has identified a pattern as an impulsivity event in two of the children, which is to raise the pencil to the level of their chin before presenting an impulsivity event and playing with the pencil. This event is not detected by smart objects or the camera. But that could be another state to apply in future work. This result reconfirms the need for this new sensor (camera) and its algorithm to advance this research.

5 Conclusions

The implementation of artificial vision as a way to manipulate the behavior of a child during the development of his/her homework provides information that the smart environment was not able to acquire at the moment. The results of the experiment and its analysis show that this new implementation for pervasive therapy for children with ADHD is necessary to identify whether the child is doing or not his/her homework. So far, there was evidence of biases in the monitored data and with this implementation we are getting more accurate and precise information. This implementation opens up the possibility of implementing the recognition of the second missing parameter, which is the change in mood of the child with ADHD during these therapy processes and also a possible act of impulsivity.

This future work will allow the environment to determine all relevant aspects of the ADHD child's behavior while doing homework and to assist the child accurately, just as a therapist would do. However, it is important to establish a secure system to protect the identity of the child because the recognition of facial expressions will be necessary during homework therapy sessions.

References

1. Manovich, L., *Computer vision, human senses, and language of art*. AI & SOCIETY, 2020: p. 1-8.
2. Voulodimos, A., et al., *Deep learning for computer vision: A brief review*. Computational intelligence neuroscience, 2018. **2018**.
3. Hassaballah, M. and A.I. Awad, *Deep learning in computer vision: principles and applications*. 2020: CRC Press.

4. Gauswami, M.H. and K.R. Trivedi. *Implementation of machine learning for gender detection using CNN on raspberry Pi platform*. in *2018 2nd International Conference on Inventive Systems and Control (ICISC)*. 2018. IEEE.
5. Irfan, M., *CNN Image classifier on Raspberry pi 3B using pre trained data*. Student of Electronics Communication, Christu Jyoti Institute of Technology Science, 2019. **8**(06): p. 14-17.
6. Hoseini, B.L., et al., *Attention deficit hyperactivity disorder (ADHD) in children: a short review and literature*. 2014.
7. DuPaul, G.J. and J.M. Langberg, *Educational impairments in children with ADHD*. 2015.
8. Naranjo, A.E., et al., *Low-Cost Assistive System for Deaf People Based on Artificial Vision*, in *Advances and Applications in Computer Science, Electronics and Industrial Engineering*. 2021, Springer. p. 249-264.
9. Berrezueta-Guzman, J., et al., *Robotic Technologies in ADHD Care: Literature Review*. 2021.
10. Peng, J., M. Debnath, and A.K.J.M.L.w.A. Biswas, *Efficacy of novel Summation-based Synergetic Artificial Neural Network in ADHD diagnosis*. 2021. **6**: p. 100120.
11. Bautista, M.A., et al., *A gesture recognition system for detecting behavioral patterns of ADHD*. 2015. **46**(1): p. 136-147.
12. Berrezueta-Guzman, J., et al., *Smart-home environment to support homework activities for children*. 2020. **8**: p. 160251-160267.
13. Dolón-Poza, M., J. Berrezueta-Guzman, and M.-L. Martín-Ruiz. *Creation of an intelligent system to support the therapy process in children with ADHD*. in *Conference on Information and Communication Technologies of Ecuador*. 2020. Springer.
14. Berrezueta-Guzman, J., et al., *Assessment of a robotic assistant for supporting homework activities of children with ADHD*. 2021. **9**: p. 93450-93465.
15. Pi, R. *Raspberry Pi Camera Module 2 NoIR*. 2022 [cited 2022 February 2022]; Available from: https://www.raspberrypi.com/products/pi-noir-camera-v2/.
16. Berrezueta-Guzman, J., et al. *Robotic assistant for the teaching in trauma accidents prevention in children of initial age*. in *2020 IEEE International Conference on Consumer Electronics (ICCE)*. 2020. IEEE.
17. López-Pérez, L., J. Berrezueta-Guzman, and M.-L. Martín-Ruiz. *Development of a home accompaniment system providing homework assistance for children with ADHD*. in *Conference on Information and Communication Technologies of Ecuador*. 2020. Springer.
18. Fujiyoshi, H., T. Hirakawa, and T. Yamashita, *Deep learning-based image recognition for autonomous driving*. IATSS research, 2019. **43**(4): p. 244-252.
19. Lee, K., et al. *Verification of normalization effects through comparison of CNN models*. in *2019 International Conference on Multimedia Analysis and Pattern Recognition (MAPR)*. 2019. IEEE.
20. Ramadhan, W., S.A. Novianty, and S.C. Setianingsih. *Sentiment analysis using multinomial logistic regression*. in *2017 International Conference on Control, Electronics, Renewable Energy and Communications (ICCREC)*. 2017. IEEE.
21. Homack, S. and C.A. Riccio, *Conners' continuous performance test (; CCPT-II)*. Journal of Attention Disorders, 2006. **9**(3): p. 556-558.