Intelligent Aided System for Autistic People

Aya Khaled
Artificial Intelligence department
Faculty of Computer and Information Science, Ain Shams University
Cairo, Egypt
aya20191703005@cis.asu.edu.eg

Alzahraa Mofed
Artificial Intelligence department
Faculty of Computer and Information Science, Ain Shams University
Cairo, Egypt
alzahraa20191703001@cis.asu.edu.eg

Amira Barakat
Artificial Intelligence department
Faculty of Computer and Information Science, Ain Shams University
Cairo, Egypt
amira20191703002@cis.asu.edu.eg

Nagwa Mohamed
Artificial Intelligence department
Faculty of Computer and Information Science, Ain Shams University
Cairo, Egypt
nagwa20191703037@cis.asu.edu.eg

Antony Nabil
Artificial Intelligence department
Faculty of Computer and Information Science, Ain Shams University
Cairo, Egypt
antony20191703004@cis.asu.edu.eg

Abstract ***Autism is a neurodevelopmental disorder that influences a person's emotions, cognition, communication, behavioral and learning patterns, their perception of their surroundings, and their social skills. It is a lifelong disability that starts when a person is born and stays with them into old age. There were existing prior applications to deal with autistic children but with how autism prevalence rate has increased in the last few years, more updated techniques are needed in terms of affordability and efficiency. We built a system that engages autistic individuals, enabling them to partake in exercises and improve their communication skills. With the help of deep learning and computer vision tasks like face recognition, facial expression, speech to text, text to speech , others and some algorithms like Haar cascade, blaze pose, CNNs, LBPH, hand landmark of media pipe and more, we designed some games in addition to educational videos that help them with concentration and movement. There is nothing about how the person looks that sets him apart from other people, meaning you can't know if someone has autism just from their appearance. So, our system doesn't have a model that recognizes if that person has autism or not. The whole system promotes involvement in meaningful activities, helping them avoid spending excessive time alone. Consequently, it assists in enhancing their condition by facilitating such interactions. We tested our system on real autistic children and it meets our design requirements which are based on effectively helping these children get better. While testing, we noticed that the system was easy to be used and games actually help children increase their concentration time span and learn to control their moves to reach more targets, we have some future recommendations to work on as well to update and improve the system.***

Keywords- neurodevelopmental disorder, autism spectrum, affordability, behavioral patterns, social skills

I. INTRODUCTION (*HEADING 1*)

*Motivation

Autism Spectrum Disorders (ASDs) are characterized by some disabilities in communication, concentration, social functioning, and movement, which can have their impact on both the children and their families. The prevalence of ASDs has potentially grown creating a need for better knowledge of the relative efficacy of various medical, behavioral, pharmacologic, and finding alternative approaches and therapies for children in addition to mechanisms for delivering services.

According to the most recent report in 2018 for the Centers for Disease control (CDC) and Prevention, one out of 44 children is diagnosed with autism in the United States, which considered as a significant increase from 2020 that estimate one out of 54 children estimated to have autism. This rate might because of increasing public awareness of autism spectrum condition and efforts to let more children have a treatment, in addition to air pollution, low birth weight and stress which contribute to raising the diagnosis.

*problem definiton

People with ASD have a wide range of abilities that differ from one person to the other. For instance, some people with autism spectrum disorders are nonverbal or have difficulties in verbal communication, others may have superior conversational abilities. Also, some people with ASD require a lot of support in their daily lives, while others may behave independently and work. Children with ASD may display certain behaviors such as not to be touched, prefer to play alone, stick to their routines, and not change them. Prior to recently, specialists discussed several forms of autism, all of which now belong under the category of autism spectrum disorder. The causes of these disorders remain unknown, only a few are known. These types are as the following: 1) Asperger's syndrome: it is no longer recognized as a distinct diagnosis but many people continue using the term to refer to themselves or other individuals who display similar characteristics, they are characterized by difficulty in social interaction and communication, although people with Asperger's syndrome tend to have average or above average intelligence, they may have issues understanding nonverbal cues or participating in reciprocal conversation. 2)childhood disintegrative disorder (CDD): known also as Heller's syndrome, it often arises in early childhood (between the age of 2 and 4 years), a severe loss of previously learned skills and abilities is noticed by children with CDD mainly in the areas of language, social interaction, and behavior. Repetitive behaviors, physical limitations and intellectual incapacity are examples of additional symptoms. 3) Kanner's syndrome, it is also known as early infantile autism or classic autism. Children with Kanner's syndrome often have serious behavioral, communicative and social interaction problems, they can hardly form a friendly connection with people, and

they are extremely sensitive to any form of sensory such as touch, sound, smell, and sight, they react violently too the surrounding environment from light, color, texture and

*Objective physical touch. So, our objective here is to

help these

children diagnosed by autism and their families and anyone responsible for them by designing a software application that will have some features some of them are educational, some are games, and others.

The second section for this paper contains an introduction showing the problem statement, motivation and the contribution made. The third section of this paper demonstrates some related work articles and reviews in the same field of using assistive technologies and intelligent aided systems for children with autism spectrum disorders. The fourth section contains a description of the used methodology with a background on each used technique. The fifth section contains the obtained results. The sixth and final section discusses the conclusion and future work.

II. Related work

This section discusses previuosly done work in the field of assistive technologies for children with autism. In Joan DiPietro [1], Arpad Kelemen [1], Yulan Liang [2], Cecilia SikLanyi [3] article "**Computer- and Robot-Assisted Therapies to Aid Social and Intellectual Functioning of Children with Autism Spectrum Disorder**" showing the importance of certain computer games and robot interventions and how they are beneficial for children with ASDs where these interventions were both reliable and cost effective. In Francisco S Melo [1], Alberto Sardinha [2], David Belo [3], Marta Couto [4], Miguel Faria [2], Anabela Farias [4], Hugo Gambôa [3], Cátia Jesus [4], Mithun Kinarullathil [5], Pedro Lima [5], Luís Luz [5], André Mateus [5], Isabel Melo [4], Plinio Moreno [5], Daniel Osório [3], Ana Paiva [2], Jhielson Pimentel [5], João Rodrigues [3], Pedro Sequeira [6], Rubén Solera-Ureña [7], Miguel Vasco [2], Manuela Veloso [8], Rodrigo Ventura [5] paper "**Project INSIDE: towards autonomous semi-unstructured human-robot social interaction in autism therapy"** they describe using mobile robots in children with ASDs therapy. They demonstrate both the software and hardware structure used to support this human- computer interaction while also mentioning the design methodology for the autonomus robust INSIDE system. Felippe Sartorato [1], Leon Przybylowski [1], Diana K Sarko [2] mentioned in their paper "**Improving therapeutic outcomes in autism spectrum disorders: Enhancing social communication and sensory processing through the use of interactive robots"** the theraputic effect that socially interactive robots have in improving social and interactive skills for autistic children and how robotic social cues integration and sensory processing can be responsible for the advantages of these social robots. Giuseppe Palestra Berardina De Carolis Sophie Sakka Thierry Le Buhé Dr. Fady Al Najjar Olivier Duris Silviu Matu Pierre Henri Bernex in their research titled "**NAO, a humanoid robot as a therapeutic mediator for young people with autism**" they mentioned the expansion of social robotics as it has the potential to exceed the limits of classic theraputic activities. It can be provided in a natural context inspite of its artificial in theraputic context it can make it possible to bring together up to four areas of expertise: engineers, teachers, pschologists, and even artists. Considering the rapid advances in this technology, this can also be used with people suffering from related disorders and slow down the development of these disorders. nature which could help in achieving progress.when a child was first introduced to NAO he showed great interest and interaction, when he sat down on the same level as the robot, that supported the idea that the child is taking part in a shared game as the NAO moves and turns its head sensing the sound and movement of the child. The NAO robot, when used.

Related Applications and Projects

No	Application name	Type of provided therapy	Used Technique
1	(NODA) Boise, Idaho: Naturalistic Observation Diagnostic Assessment	Remote Diagnosis. Used by parents by uploading videos of their children for observation and for categorizing their children's behaviors	Novel imaging technology
2	Manatee, Denver: Bibli robot	Determining challenging issues for children with autism spectrum disorders where it collects information and sends it to therapists	Facial expressions and gestures
3	Starfall mobile application	Helps in making the child ready for school as it includes coloring, reading, songs and games therefore improving the child's skills	Touch screen and picture based
4	LAMP words for 4life	A therapeutic approach based on neurological and motor learning principles that contributes in building speech and language skills	Motor Planning (LAMP™) approach. It is symbol based paired with text

III. Methodology and materials

This section will present the different techniques and methods that were used in our system along with a brief description about each one including face recognition, facial expressions, full body estimation and poses detection, and games like card game, body spinning ball, snake game, car game, virtual paint, and programs like alphabet teaching program, numbers teaching program, colors teaching program, prayer teaching program, and human emotions teaching program, ablution teaching program.

Our system is supposed to be aiding the autistic children through educational games specified to them (not any games) that allow them to do some exercises that will maintain and improve their concentration and communication skills.

A. Face recognition

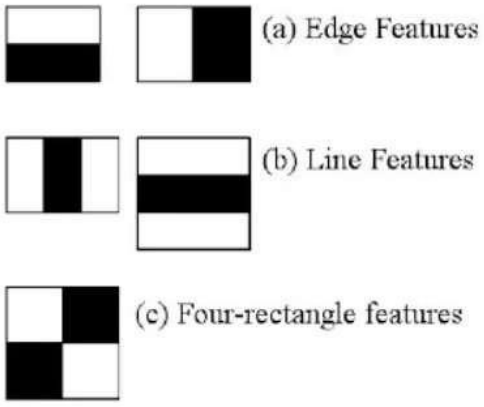
In this part , at first the child has to choose between two things either to go through the application altogether or to choose by himself what he want, if he chooses to go through the application, here we will need the help of facial expression recognition(will be discussed later) to recognize the behavior of the child and then based on it a pre-determined sequence of games and videos will start,

otherwise if he chooses by himself what he want there will be two options either to learn through educational videos or to play some games. The algorithms and techniques used in this section are Haar cascade classifier and Local Binary Pattern (LBP) and Histogram Oriented Gradients (HOG), LBPH face recognizer.

1) Haar cascade classifier

It is an object detection algorithm used to identify faces in an image or a real time video, the algorithm uses edge or line detection features proposed by viola and jones in their research paper "Rapid Object Detection using a Boosted cascade of simple features" published in 2001. It is a machine learning-based approach in which a lot of positive images (images that we want the classifier to identify) and negative images (images of everything else that doesn't contain the object to be identified) are used to train the classifier. This classifier is composed of four stages:

1.1) Calculating Haar features: features may be edges, lines, or four-rectangle features, collecting these features considered as the first step in the algorithm at which some calculations are performed on adjacent rectangular regions at a specific location in a detection window, that involves summing the pixel intensities in each region and calculating the differences between the sums.



Types of Haar features. (Image Source)

2.1) Creating integral images: From the previous step, instead of computing at every pixel, integral images are used, it creates sub-rectangles and creates array references for each of those sub-rectangles, they are used later to compute the Haar features.

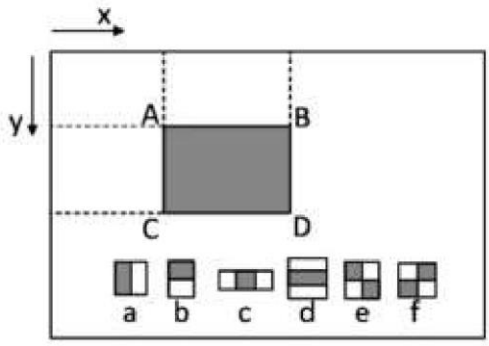


Illustration for how an integral image works. (Image Source)

Thus, the integral images are computed as follows, it is generated where each pixel's value represents the sum of all the pixels located to its left and above in the original image. Ultimately, the sum of all the pixels in the original image is represented by the last pixel located at the bottom right corner of the integral image.

3.1) Using Adaboost: Some of the Haar features obtained in the two previous steps will be irrelevant when doing object detection, so the best features must be determined, and that is the goal of this step. Adaboost is a machine learning algorithm that selects the most important and relevant features in the input data and trains weak classifiers to use these features, the algorithm combines the predictions of these weak classifiers to create a strong classifier that can accurately detect the objects. The weak learners(classifiers) are created at which a window is moved over the input image and Haar features are computed for each subsection of the image, then the difference between the two rectangular regions within each subsection is evaluated and compared with a learned threshold that distinguishes objects from non-objects. Because these are "weak classifiers," a large number of Haar features is needed for accuracy to form a strong classifier.

4.1) Implementing cascading Classifiers: The last step is to combine all these weak classifiers into a strong learner(classifier) using cascading classifier. The cascade classifier consists of multiple stages, and each stage is composed of a set of weak learners(classifiers). Boosting is used to train these weak learners(classifiers), which enables the creation of an accurate classifier by taking the mean prediction of all weak learners. And based on this prediction, the classifier decides whether to indicate the presence of an object (positive) or move on to the next region (negative). The stages are designed to reject the negative samples since the majority of windows do not contain relevant objects.

2)LBPH face recognizer

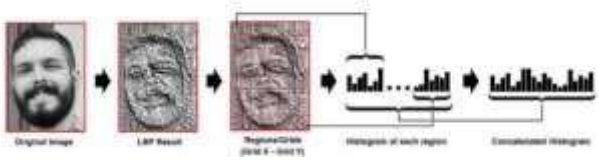
It is a face recognition algorithm based on a local binary operator which is designed to recognize both the side and front face of a human. It uses four parameters the radius, neighbors, Grid X and Grid Y. This algorithm is composed of five stages as the following:

2.1) Training Facial Images Dataset: It is the first step in the algorithm at which it is trained on facial images of people to be recognized later and setting an ID for each image.

2.2) Applying the LBP Operation: At first an intermediate image is created that describes the original image in a more better way that is by highlighting the facial characteristics of the input images, this is done by using the concept of a sliding

window based on these two parameters the radius and neighbors, so for example if we have a facial gray-scale image, the sliding window can be of any size we choose, it may be 3x3, 5x5 or 7x7, let's consider it in this example a 3x3 sliding window, where each 3x3 matrix contains the intensity of each pixel (0~255), then take the central value of the matrix to be used as a threshold, then if the value of the neighbor is greater than or equal to the central value it is set as to one otherwise it is set to zero. Thus, a total of 8 binary values are obtained from the 8 neighbors, after combining these values we get 8-bit binary number which is translated to decimal number, this decimal number is called the pixel LBP value.

2.3) Extracting the Histograms: In this step, Grid X and Grid Y parameters can be used to divide the image into multiple grids as shown here



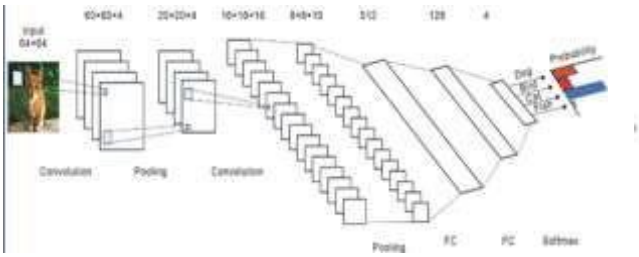
2.4) Concatenating the Histograms: Then, we need to concatenate each histogram to create a new and bigger histogram. Supposing we have 8x8 grids, we will have 8x8x256=16.384 positions in the final histogram. The final histogram represents the characteristics of the image original image.

2.5) Performing Face Recognition: And finally, the algorithm is already trained in this step and each histogram created is used to represent each image from the training dataset, so to find the image that matches the input image, simply a comparison is made the histograms of the two compared images and return the image with the closest histogram. This is done by using any similarity measure like Euclidean distance, chi-square, absolute value.

The algorithm will output the ID of the image with the closest histogram, the calculated distance which can be used as a confidence measurement, and the lower confidence are the better ones as it represents the smallest distance between the two compared histograms.

B. Facial expressions

This section in our system is used when the child uses "let's play", at which several images of the child are taken which



In our system we designed a CNN for fa cial expression recognition using Keres deep learning library and the

are then used to determine whether he is happy, sad, angry, neutral, or surprised. The algorithms and techniques used in this section are Convolutional Neural Networks (CNNs) and Haar Cascade classifier. The Haar cascade classifier was discussed briefly above in "Face Recognition techniques and algorithms", so in the following will be w brief about CNNs.

1) Convolutional Neural networks (CNNs) A Convolutional Neural Network is one of the main categories in image recognition, image classification and object detection, it is a class of Neural Networks that is specialized in processing large data like images which consists of more and more pixels according to its resolution. CNN consists of input, output and as well as multiple hidden layers. The hidden layers of a CNN typically consist of convolution layers, pooling layers, fully connected layers, and normalization layers.

ImageDatagenerator class to preprocess and augment the image data. The architecture consists of several convolutional layers, each followed by an activation function, batch normalization, and max pooling. The final layers are fully connected layers that perform classification. The model is trained using the categorical cross-entropy loss function and the Adam optimizer, and to prevent the overfitting, the code uses early stopping, model checkpointing, and learning rate reduction techniques. The model is trained for 25 epochs, and the performance is evaluated using accuracy as the metric.

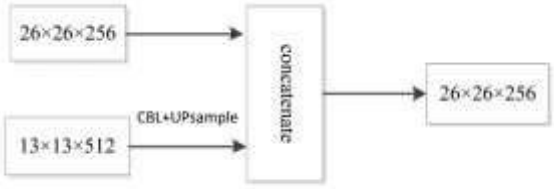
C. Car Game

In this section, the algorithm used is "BlazeFace".

BlazeFace is a machine learning model originally wa for Mediapipe provided by Google to rapidly detect the location and keypoints of faces. The position of the face and the keypoints if the face can be obtained simultaneously, there are six keypoints as follows: eyes, nose, ears, and mouth. It is also possible to detect multiple people at the same time. The architecture of the model is designed to perform very fast on mobile GPUs. The architecture is as following: it uses an improved network based on MobileNet, the efficiency of 3x3 depthwise convolution on a 56x56x128 tensor is evident from its execution time of 0.07ms on the iPhoneX. However, the subsequent 1x1 convolution that converts 128 channels to another 128 channels takes 4.3 times longer at 0.3ms. This indicates that increasing the kernel size for the depthwise convolution is a cost-effective measure. The authors thus suggest substituting the 3x3 depthwise convolution with a 5x5 depthwise convolution, which reduces the model's depth and speeds up the process.

D. Card Game

In this section, the algorithm used is tiny-YOLOv3 (You Only Look Once, Version 3), and tiny-YOLOv3 is a simplified version of YOLOv3, which has a smaller number of convolution layers than YOLOv3 which means that into 26 × 26 × 256. The feature map of 26 × 26 also is taken from the earlier in the network and merged with the



tinyYOLOv3 doesn't need to occupy a large amount of memory thus reducing the need for the hardware. Overall, YOLO is a real-time object detection algorithm that identifies specific objects in videos, live feeds, or images, YOLO machine learning algorithm uses features learned by a deep convolutional neural network to detect an object. The architecture of tiny-YOLOv3 is as follows: the network has seven convolution layers with 3x3 convolution kernels and

one convolution layer with 1x1 convolution kernel, six layers of maxpooling are used to reduce the parameters, the object is predicted by using a two-scale prediction network with the output feature map of 13x13 and 26x26. It uses the upsampling to extract the features. In the following figure, the 13 × 13 feature map passes the convolution layer and upsampling layer, this turns the 13 × 13 × 512 feature map upsampling feature by concatenation. Finally, the output feature map of 26 × 26 is formed.

E. Snake Game

In this section, the algorithm used is Hand landmark detector of cvzone. HandLandmarker is a computer vision model that is designed to detect and localize key points on a human hand in real-time. The model is based on deep learning and uses a convolutional neural network (CNN) architecture to analyze images and identify the location of various landmarks on the hand. The model is designed to work in real-time, which makes it suitable for applications such as gesture recognition, virtual and augmented reality, and robotics.

F. Spin Game

In this section, the algorithm used is Blaze Pose. It is a human pose estimation model developed by Google that uses machine learning to estimate the 2D and 3D poses of a person from a single RGB image or video frame, it is part of the MediaPipe framework provides a pre-trained. It consists of two machine learning models: a Detector and an Estimator. The architecture is as the following: The detector is an SingleShot Detector(SSD) based architecture, given an input image (1,224,224,3), it outputs a bounding box (1,2254,12) and a confidence score (1,2254,1)., the 12 elements of the bounding box are of the form (x,y,w,h,kp1x,kp1y,…,kp4x,kp4y), where kp1x to kp4y are additional keypoints. While the Estimator uses heatmap for training but computes keypoints directly without heatmap for faster inference. It outputs a set of 195 landmarks with dimensions of (1,195), where each landmark consists of 5 values (x, y, z, visibility, and presence)., the 50 landmarks represent 33 key body parts, and for each body part, there are 5 landmarks representing different aspects of the body part's position and appearance.

IV. Experimental and Results

The face recognition system in this system is trained on a dataset consisting of 30 images for each user. The system's accuracy is measured using a confidence score, which is calculated based on the distance between histograms. A threshold value of 100 is used to differentiate between known and unknown users.

we utilized a dataset obtained from a kaggle competition to train a facial expression recognition system. To improve the overall accuracy of the system, the dataset was split into training and validation sets. The validation accuracy obtained was 68.19%

The developed system is aimed at providing an educational and entertaining experience for autistic children. In order to evaluate the system's performance, a 5-year-old autistic child diagnosed with ADHD was selected as an enduser. The system was found to run smoothly with all features implemented and working well.

During the evaluation, the child exhibited a positive response to the system's colorful GUI and was fully engaged throughout the interaction. The child watched educational videos with focus and interacted with the questions at the end of the videos. Although the spinball and car games presented some initial challenges, the child learned how to play after practicing for a while. The snake game and paint game were particularly enjoyable, especially the ability to change the drawing color while drawing.

Two different sequences were tested, "happy" and "neutral", which determined the order of the games and educational videos. The child was responsive to both sequences, and the system was found to be successful in educating and entertaining the child. Overall, the system demonstrated its potential for providing an engaging and beneficial experience for autistic children.

V. Conclusion and Future Work

The paper presents an interactive educational system designed to help children with Autism Spectrum Disorder (ASD) manage their symptoms in a fun and engaging way. The system includes face recognition and facial expression recognition modules, implemented using the Local Binary Patterns Histogram (LBPH) recognizer and a custom Convolutional Neural Network (CNN) respectively. The LBPH recognizer uses a confidence score as an accuracy metric, while the CNN model is trained using the Keras deep learning library and the ImageDataGenerator class for image preprocessing and augmentation. To prevent overfitting, the code employs early stopping, model checkpointing, and learning rate reduction techniques, and the model is trained for 25 epochs, with accuracy as the evaluation metric. The dataset is split into training and validation sets to improve accuracy, and the proposed method achieves a validation accuracy of 68.19%.

The system also includes educational videos on various topics and several interactive games, such as a car game, snake game, spinball game, virtual game, and card game. The games are played using hand and head movements, aiming to improve motor functions and communication skills. The system is tested on a child with ASD and ADHD, and the results show that the child responded well to all features. The child's performance and grasp of the concepts improved over the testing time, indicating that the system succeeded in achieving the desired benefits.

Overall, the paper presents an interactive system that combines face recognition, facial expression recognition, educational videos, and interactive games to provide an engaging and beneficial experience for children with ASD. The proposed method achieves good accuracy, and the system is found to be effective in helping children manage their symptoms and improve their motor functions and communication skills. The proposed interactive system has the potential to assist children with ASD in adapting to their disorder and managing its symptoms. The system aims to improve their communication skills and motor functions over time. Additionally, the system can be integrated into a robot or an online server and distributed on a wider scale to reach more children with ASD. To further enhance the system, the educational videos can be transformed into interactive programs to fully leverage their benefits.

To ensure the effectiveness of the system, thorough testing should be conducted on a large number of end-users, and any necessary modifications should be made. Furthermore, it is crucial to test each sequence of games and

videos for different modes and adjust them accordingly to ensure optimal results.

These measures will help to improve the system's efficacy and provide a better experience for children with ASD.

ACKNOWLEDGMENT

We offer our sincerest gratitude to our supervisors, Prof. Dr. Abeer Mahmoud, and T.A. Andrew Magdy, who have supported us throughout our thesis with their patience, knowledge and experience.

REFERENCES

[1] https://www.illinoislend.org/new-blog/2023/1/20/a-brief-insight-intoautism-diagnosis-and-treatment-by-christy-yoon-special-educationtrainee, January 20,2023

[2] https://medium.com/analytics-vidhya/haar-cascades-explained-38210e57970d , December 21, 2020

[3] https://medium.com/@RaghavPrabhu/understanding-ofconvolutional-neural-network-cnn-deep-learning-99760835f148 , march 4, 2018

[4] https://medium.com/axinc-ai/blazeface-a-machine-learning-modelfor-fast-detection-of-face-positions-and-key-points-5dcfb9429d72 , may 18, 2021

[5] AashalKamdar, "Haar Cascades for object Deetection",https://www.geeksforgeeks.org/python-haar-cascadesfor-object-detection/

[6] Valentinaalto, July 16, 2019, "Haar Cascade: Integral Images", https://datasciencechalktalk.wordpress.com/2019/07/16/haar-cascadeintegral-image/ [7] Kelvin Salton do Prado, November 10, 2017, "Face Recognition: Understanding LBPH", https://towardsdatascience.com/facerecognition-how-lbph-works-90ec258c3d6b

[8] Girija Shankar Behera, December 24, 2020, "Face Detection with Haar cascade", https://towardsdatascience.com/face-detection-with-haarcascade-727f68dafd08

[9] Paul Viola, Michael Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features", https://www.cs.cmu.edu/~efros/courses/LBMV07/Papers/viola-cvpr01.pdf

[10] Wendong Gai, Yakun Liu, Jing Zhang, Gang Jing, march 23, 2021, "An improved Tiny YOLOv3 for real-time object detection", https://www.tandfonline.com/doi/full/10.1080/21642583.2021.1901156

[11] Swati Sucharita Barik, Sasmita Nayak, Stitiprajna Panda, Aeisuriya Tripathy, march 2020, "Human Face Recognition using LBPH",

[12] https://www.medic8.com/mental-health/autism/classic-autism.html

[13] Sara Karim, Nazina Akter, Muhammed J. A. Patwary, Md.Rashedul Islam, November 2021, "A Review on Predicting Autism Spectrum Disorder (ASD) meltdown using Machine Learning Algorithms",

[14] https://www.researchgate.net/publication/357727695_A_Review_on_Predicting_Autism_Spectrum_DisorderASD_meltdown_using_Machine_Learning_Algorithms

[15] Ernst Vanbergeijk, Oren Shtayermman, October 2005, "Asperger's Syndrome", https://www.researchgate.net/publication/272119974_Asperger's_Syndrome

[16] https://www.researchgate.net/publication/342987802_Human_Face_Recognition_using_LBPH