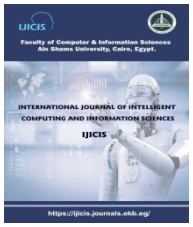
IJICIS, Vol.20, No.2, 89-111
DOI: 10.21608/ijicis.2020.46360.1034

International Journal of Intelligent Computing and Information Sciences

https://ijicis.journals.ekb.eg/



COMPUTER VISION TECHNIQUES FOR AUTISM SYMPTOMS DETECTION AND RECOGNITION: A SURVEY

Esraa T.Sadek*
Department of Computer systems, Faculty of Computer and Information Sciences,Ain Shams University
Cairo, Egypt
esraa.sadek@cis.asu.edu.eg

Noha A. Seada
Department of Computer systems, Faculty of Computer and Information Sciences,Ain Shams University
Cairo, Egypt
noha_sabour@cis.asu.edu.eg

Said Ghoniemy
Department of Computer systems, Faculty of Computer and Information Sciences,Ain Shams University
Cairo, Egypt
ghoniemy1@cis.asu.edu.eg

Received 2020-10-14; Revised 2020-11-26; Accepted 2020-12-05

Abstract: *Autism spectrum disorder (ASD) is a world-threatening mental developing disorders that recently appeared widely, due to its diagnosis complexity as well as lack of evidence of its real causes. Many researchers have afforded great effort to precisely identify this syndrome and its symptoms. This survey provides a comprehensive study of autism spectrum disorder, its types, symptoms, prevalence, and developments in its diagnosing. Six categories for autism exposure and identification are currently investigated; clinical monitoring, genetics and blood analysis, Functional magnetic resonance imaging (fMRI), Electroencephalography (EEG) based investigation, wearable sensors and finally computer vision-based techniques. Computational technologies, especially computer-vision, machine learning and neural networks techniques have added great advances in detecting autism and these techniques are comprehensively reviewed in this paper. Also, medical assisting computer vision-based framework is proposed to detect observable autism symptoms. The proposed framework utilises recent and efficient techniques that can be used to produce accurate diagnosing results.*

Keywords: *ASD, Autism spectrum disorder, autistic symptoms detection, autism signs detection repetitive motor behaviors, autistic self-stimulatory or stereotypy behaviors, Activity Recognition and Classification, Computer Vision*

* Corresponding author: Esraa T.Sadek
Department of Computer systems, Faculty of Computer and Information Sciences, Ain Shams University, Cairo, Egypt
E-mail address: esraa.sadek@cis.asu.edu.eg

1. Introduction

Autism Spectrum Disorder (ASD) is a complicated mental developmental syndrome affecting children aging from 2 to 5 years [1] [2]. ASD is a set of defects, with varying degrees, in socialization, communication, expressing and understanding emotions and stereotype. Autistic children might perform weird behaviors in several forms, like getting nervous by minor changes, having strong attachments to possessions, underestimate of danger, repeating words, or phrases over and over, evading eye interaction and having tendency to be alone. Regrettably, autism is a lifelong disorder. However, research demonstrate that early mediation treatment administrations can significantly enhance a youngster's development.

Several reasons encourage researchers to contribute to autism diagnosing research. One of them is that there is no proven cause of autism till now; however, research indicate that genetic defects are the expected cause. The second reason is the intensive rise in autism population. One of each 110 children is diagnosed with autism [3], and 1 of each 59 is expected to be autistic, based on the Centers for Disease Control and Prevention (CDC) latest report [4], Figure 1.a, b. The third reason is that autism population dramatically exceeds the number of clinicals who can diagnose autism. As a result, the waiting list for an assessment at the main Autism Spectrum Disorder Clinic at the University of Minnesota is a half year for kids aging 4 and under [5]. Finally, autism early detection can have an effective impact on treatment in which children can return to a point as they were typically developing. Males are expected to be diagnosed with autism five to six times more than females [6]. As the reason of autism is still unknown, researchers expected that there are some defections in genetics causing autism. However, detecting autism using genetic analysis is still unclear, so clinicals usually diagnose autism by monitoring child's behaviors [3]. Autism diagnosis has been improved during the last decade as its symptoms are being more conspicuous [7]. Current detection approaches include detection using genetics and blood analysis, EEG-based investigation, Functional Magnetic Resonance imaging (fMRI) based on blood-oxygen-level dependent (BOLD) techniques, clinicals observation only, wearables and sensors and computer vision techniques. Each category has massive research and efforts which will be overviewed in the next sections.

Each of these approaches has its own limitations. However, applying computer vision techniques to recognize ASD has many benefits such as reducing clinical expenses in detecting autism and expanding standard tools to detect autism faster and easier. It is worth noting that, proposed computer vision programs are not supposed to suppress medical professional's work.

This paper is sorted out as follows. Section II summarizes the classes of autism spectrum, categorizes of autism symptoms, and approaches for detecting autism. Section III reviews Computer vision based human activity recognition, evaluation metrics, human activity recognition datasets, as well as the challenges and research objectives for autism discovery using computer vision. Section IV concludes the paper.

2. Autism signs detection and recognition

Classes of Autism Spectrum

Autism Spectrum Disorder has three classes: Autistic Disorder (classical autism), Asperger Syndrome and Pervasive Developmental Disorder Not Otherwise Specified (PDD – NOS).

a. Autistic Disorder (Classic Disorder)

The most known type of autism. There is normally nothing about how autistic children look like to be differentiated from other individuals. Yet they may convey, connect, carry on, and learn in manners that are not the same as other normal children. Classic autism's symptoms appear in some mental defects, language delays, strange behaviors, and social and communicational difficulties .

b. Asperger Syndrome

Although indications are available early throughout everyday life, Asperger disorder is normally analyzed when a youngster is school aged. Similarly, as with different ASDs, researchers do not know precisely what causes Asperger disorder, but it is realized that the mind of somebody with this condition works uniquely compared with of somebody without Asperger disorder. This type of autism is less in severity than classic autism. The symptoms of Asperger syndrome are almost close to classic autism; however, no language or mental disabilities appear here. Asperger Syndrome 's symptoms include having trouble in understanding emotions and body language, avert eye contact, prefer to be alone and sometimes like to cooperate with others but do not know how to do so and showing limited and sometimes strange interests .

c. Pervasive Developmental Disorder Not Otherwise Specified (PDD-NOS)

Individuals who have some of the symptoms of classic autism or Asperger syndrome, yet not all, might be determined to have PDD-NOS. Individuals of this type, as a rule have less indications than those with medically Autistic disorder. The symptoms are just social and communication challenges.

Categorizes of autism symptoms

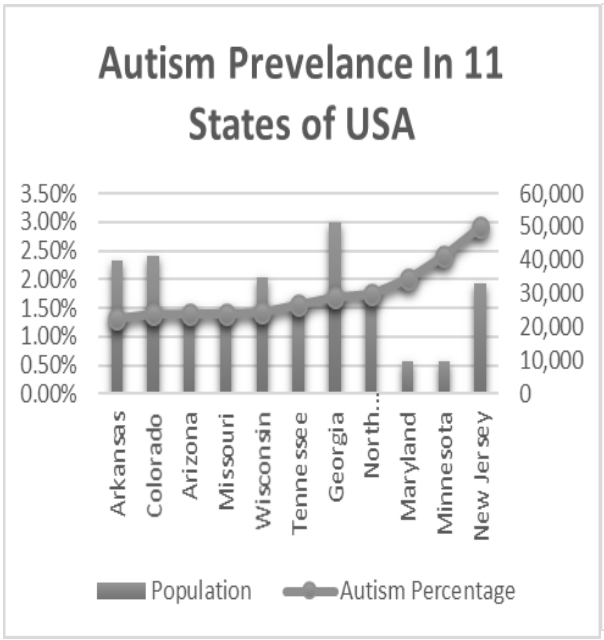


Figure 1.a Autism prevelance in 11 states of USA

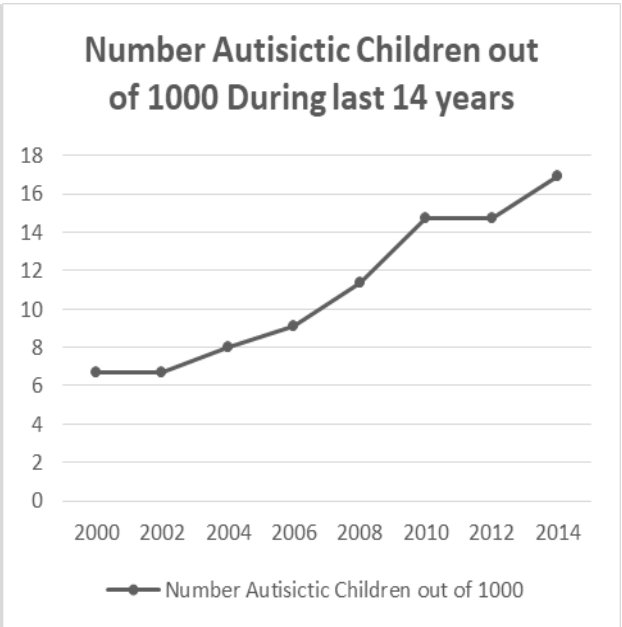


Figure 1.b Number of autistic children during the last 14 years

Although that autism cause is mysterious and there are no special signs that differentiate autistic children in look from well-developed ones, the difference appears in communication, socialization, behaving and learning. The learning, understanding, thinking and problem-solving skills of autistic children can vary from talented to very challenged. Some require a special deal of assistance in their every-day lives; others require less. The autism symptoms can be categorized into visual symptoms, social and communication symptoms, and developmental symptoms.

a) Visual Symptoms

One obvious autism symptom of Autistic children is eye contact averting. Visual behaviors are important indicators of autism which can lead to early diagnosis and treatment in children. Several research [8]– [12] have conducted their research to prove the usability of analyzing visual behaviors for early autism detection in children during early stages. Zawigenbaum et al. [9] [10] claimed that special signs of visual attention, difficulties in disengagement and attention can be detected during the first year of children. As a subfield of visual behaviors, eye tracking is one of early symptoms of developmental disorders in autistic children. According to studies in [13], eye tracking has been found to be effective indicator of autism in new-born and toddlers especially because it does not require motor or language response. Lack of eye contact reduced looking time to faces, and disengagement of attention can be detected by several ways starting from daily monitoring up to sophisticated computer vision systems.

b) Social and communicational Symptoms

Autistic children usually have low tendency to initiate communication or share interest and they also have weak emotional understanding. They usually interact differently from well-developed children. Communication skills do not develop naturally enough in about 30% to 50% of autistic children.[14]

c) Developmental Symptoms

Autistic children usually do not develop in a normal rate. Esposito [10] said that motor development can reveal autism biomarkers. Atypical motor patterns like irregular gait or toe walking are considered as clear early autism indicator [15]. However, autistics' mental skills might grow faster than their social and language skills. Also, autistic children may have higher ability to memorize vocabulary while unable to run properly. Observing atypical behaviors as well as delay in the typical developing symptoms are predicators of autism. Researchers [1], [16]– [19] recorded some red flags of delay of typical development signs like pointing to or showing objects, emotional understanding, imitation and deficiency of attention .

d) Repetitive behaviors

Autistic people can have one or several types of monotonous or constricted behaviors. According to Repetitive Behavior Scale-Revised (RBS-R) [20], repetitive or restricted behaviors have six categories:

i. Stereotyped behaviors (Self-stimulatory): Monotonous movements, for example, hand fluttering, head rolling, or body shaking.
ii. Compulsive behaviors: Time-expending practices proposed to diminish tension that an individual feel. Compulsive behaviors constrained to be performed more than once, for example, putting in objects in a particular form, examining items, or hand cleaning.
iii. Monotony: Autistics have a minimal tendency to change; for example, they feel anxious with shifting or moving furniture or painting color. Obliging them to change their atmosphere make them nervous .
iv. Ritualistic behavior: Constant case of daily habits, for example, a perpetual menu or a dressing routine. Ritualistic behavior is nearly connected with sameness and an autonomous approval has proposed combining the two factors [20]. Although, normal children tend to regularly change their routines, ways of playing and movements, autistic children face difficulties in changing their routine [21]. They can repeat their activities hundreds of times daily without getting bored .

v. Restricted interests: abnormal attachment to objects or interests, such as fixation to specific television show, toy, game, or dress.
vi. Self-hurt: Behaviors like eye-jabbing, skin-picking, head-banging, and hand-gnawing.

Among these repetitive behaviors that reveal autism disorder in young children; self- stimulatory or stereotypy behavior, which is a repeated movement of body parts or items [22], is one of the significant indicators of autism. Stereotypical behaviors limit the development of learning and social abilities in autistics [23]. Several researchers [24]– [27] stated that autism signs significantly vary with age. Stereotypy is a critical feature of autism in children of approximately 2-5 years old [1], [2] but unfortunately cannot be considered as autism sign in infancy [28]. Although that repetitive motor behaviors are important symptoms of autistic subject and is performed daily, clinicals cannot keep monitoring autistics all the time. Accordingly, these behaviors can be recognized from recorded videos especially because it almost follows some pattern; however, it is not straightforward process specifically in uncontrolled environment. Researchers in [22] have conducted and annotated a dataset (SSBD) of self-stimulatory behaviors videos in children to be used in further research. The dataset contains three stereotypy behavior which are arm fluttering, head banging and body rolling. They state that the automated investigation of the videos is challenging due to variety of recorded situations.

3. Approaches for detecting autism

As autism detection using genetic analysis is still unclear, clinicals prefer to diagnose autism by monitoring child's behaviors [3]. Traditional ways of detecting autism in children is based on monitoring them and their behaviors and comparing by the developmental references tools like SMMD (Statistical Manual of Mental Disorders), the Autism Diagnostic Observation Schedule (ADOS) [29], the Autism Diagnostic Interview [30], the Autism Observation Scale for Infants (AOSI) [31], Autism Spectrum Scan Questionnaire (ASSQ) [32], Children Asperger Syndrome Test (CAST) [33] and the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) [34], to correctly scale the disorder, which is time and effort consuming. The Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5, 2013) aims to diagnose two autism indicators: (a) social and communicational delays (b) up-normal repetitive behaviors like SMM. The common forms of SMM which usually appear in different levels of severity in autistic children are hand shaking, body rocking, head banging and finger flipping. According to the survey proposed by A. A. Baumeister et al. [35], about 60% up to 10% of autistic people have some sort of these symptoms. The average age for detecting autism is 5 years old but earlier detection can lead to better treatment [36]. Several studies [24] [1] [37] have proved that autism has become easier in detection due to its high prevalence. There are several autism diagnosis techniques, including genetics and blood analysis, EEG based investigation, Functional magnetic resonance imaging (fMRI) based on blood-oxygen-level dependent (BOLD) techniques, clinicals observation only, wearables and sensors and computer vision techniques .

a) Approaches for detecting autism using genetics and blood analysis

There have been some researchers applying genetic sciences to identify autistic youngsters [38]– [41]. Shen et al. [38] announced proof recommending that hereditary variables have a strong relation for causing ASD. They propose a clinical hereditary test, which incorporates Karyotype, delicate X testing, and CMA all together. Wang et al. [39] in their research found a robust link between specific genes and being susceptible to be diagnosed by ASD. Also, Veenstra-VanderWeele et al. [41] published a survey

article studying genetics oddities effect in ASD, and Vorstman et al. [40] clarified irregularities of autistic and schizophrenic children genes. Momani et al. [42] proposed another medical blood-based indicator that can be utilized to detect ASD. They analyze ASDs via investigating the blood plasma. They revealed that the peptide example of kids with a mental imbalance is unique in relation to non-extremely introverted ones. They additionally examined the connection between the side effects of ASDs and distinctive peptide designs. In their investigation, an examination that can describe the peptide utilizing just 3-ml of blood is proposed. After effects of the examination of the blood tests of 18 youngsters with a mental imbalance and 30 non-extremely introverted ones between 3– 12 years of age exhibited 86% affectability and 77% specificity. Different specialists [43] focused their work on Fragile X Syndrome (FXS) which is the very well-known genetic reason of autism [44]. FXS's population reaches about 100,000 persons in the United states. People with FXS have the same popular symptoms of autism.

b) Approaches for detecting autism using EEG-based investigation

Brain signals are used recently to reveal many medical secrets. Several researches used Electroencephalography as a medical biomarker for autism detection and diagnosis [45]– [47]. William Bosl et al. [47] claim that mental developmental disorders like autism show brain variations before behavioral symptoms show up. They proposed an approach that studies EEG signals of children and finds any strange variations and if found is used as red flag indicating high probability of being diagnosed with autism. They applied their proposed approach on a group of typically developing toddlers and others with high probability of being diagnosed with autism aging from 6 to 24 months. Their approach succeeds to achieve 80% accuracy rate of correct diagnosis for a set of children aged from 9 to 12 months. Also, Kamel et al. [46] has used EEG signals to detect autism. Kamel et al. applied Fourier Transform (FFT) and Related Fisher Linear Discriminant (RFLD) on a dataset of 15 children aged from 10 to 11 and he achieved 92% accuracy rate. Like previous research, Sheikhani et al. [45] utilized Quantitative EEG (QEEG) signals to distinguish between autistic and non-autistic children. He applied statistical methods on a group of autistic and non-autistic pupils ageing from 6 to 11 accuracy rate up to 96.4% .

c) Approaches for detecting autism using fMRI

Functional magnetic resonance imaging (fMRI) is a new technique relies on blood-oxygen-level dependent (BOLD) methods. FMRI has been used to monitor brain behaviors during mental and functional activities. FMRI has two brain provocations which are task fMRI (tfMRI) and task-free or sometimes called resting fMRI (rsfMRI). Brain imaging techniques especially FMRI have been effective way of diagnosing autism as they provide effective brain deficiency indicators. FMRI inducements can be represented in visual, auditory or any form of behavior representation. Usually, researchers incorporate the two classes of FMRI to achieve more reliable diagnosis results. Guillaume Chanel et al. [48] have proposed a framework relying on Multivariate pattern analysis (MVPA) of FMRI readings gained from two tests. Guillaume's proposed framework utilized two classification algorithms which are Support Vector Machines (SVMs) and Recursive Feature Elimination (RFE) to differentiate between autistic and typically developing subjects. Guillaume's framework gained accuracy ranged from between 69% and 92.3% indicating the effectiveness of cooperating brain imaging and machine learning techniques. Also, N. C. Dvornek et al. [49], proposed a recurrent neural network classification solution that distinguish between autistics and typical control subjects based on their FMRI readings. Dvornek's achieved classification results ranged from 51.8% to 69.8% .

To extract the useful information within the medical form of FRMI readings, several computational approaches have been suggested, like general linear model (GLM) for tfMRI, independent component analysis (ICA) for rsfMRI, as well as many other approaches like wavelet al algorithms, Markov random field (MRF) models, mixture models, autoregressive spatial models, Bayesian approaches, etc.

In these approaches, GLM is commonly used technique because of its efficiency, easiness, strength, and wide accessibility .

d) Approaches for detecting autism using wearables sensors

Several researches have been conducted to automatically detect autism and help clinicals to examine autistic kids by utilizing wearable sensors. F. Albinali et al. [50] and M. S. Goodwin et al. [51] have utilized 3D accelerometers to collect behavioral patterns data and to identify arm flapping and body rocking behaviors. The used accelerometers were placed on each wrist and chest. Although that the produced results were promising but restricting autistic child to wear sensors increase their feeling of anxiety. Westeyn et al. [52] also proposed an accelerometer-based framework of hand flapping detection. Westeyn et al. [52] achieved accuracy of 69% using Hidden Markov Models applied on a dataset of healthy individuals mimicking hand flapping.

Also, Min et al. [53]– [56] constructed a dataset of 40 hours reading of 3-axis accelerometers placed on autistic subjects. After that, Min et al. [53]– [56] employed some semi-supervised classification methods on the repetitive motor behaviors features. The proposed model achieved accuracy of 86% and 95% for hand flapping and body rocking, respectively. Min and Tewfik [54] developed a technique to automatically detect repetitive and self-injuries activities by analyzing readings of wearable sensors on 4 children. Min et al. [54] got the benefits of used Linear Prediction Coding (LPC) to classify repetitive movements. Min et al. achieved high rates of accuracy in detecting self-injurious behaviors, flapping, and rocking by 95.5%, 93.5%, and 95.5% respectively. The proposed framework by Min et al. was tested on many children aged from 8 to 48 months and divided into three groups. Typically developed group of 106 children, language delayed group of 49 children and 77 children in autistic group. Also, Goodwin et al. [57]– [59] recorded 3-axis accelerometers data worn by 6 autistic children in school rooms. Goodwin et al. used time and frequency-domain features in their research. As the use of recurrence plot has added a lot to autism research, Goodwin utilized recurrence plot to solve autism mystery achieving accuracy of 81.3% in detecting SMM. Also, Fusaroli et al. [60] and Romero et al. [61] have applied recurrence plots to evaluate social interaction. Romero et al. [61] used recurrence plots to study social motor coordination, and Bhat et al. [62] in diagnosing autism using Electroencephalography. U. Großekathöfer et al. [63] have introduced new selection of recurrence plot-based features. U. Großekathöfer et al. [63] have applied wearables quantification analysis in cooperation with classification algorithms to maximize classification accuracy. to get higher accuracy. U. Großekathöfer et al. [63] have achieved up to 9% increase in accuracy by applying these new set of features on Goodwin's dataset .

Despite the promising results of proposed approaches, the major drawback of them is the force of carrying the sensors by autistic children that increase their feeling of anxiousness. In any case, depending on accelerometers in detecting developmental disorders' symptoms have numerous restrictions appeared in all publications. First, researchers characterize SMM behaviors in terms of accelerometer frequency, statistical features, or movement of joints in different axes. Usually, these features differ over time and among individuals. Therefore, these features are hardly applied in dynamic SMM recognition. Second, the utilized classification algorithms apply complex permutations on the extracted features which are computationally expensive. Third, using multiple sensors types on autistic children results in multiple various detection of the same SMM activity.

To overcome the limitation of the difference of behavioral patterns inter and intrasubject, N. M. Rad et al. [23] have proposed using deep learning to learn distinguishing features for SMM pattern recognition. Features learning in cooperation with transfer learning techniques applied in convolutional neural network CNN provides accurate SMM detection. Features learning techniques generalize behavioral recognition approaches, so it can be applied on any new dataset. N. M. Rad's results showed the usability of CNN which provided results close to traditional classification techniques. Providentially,

variation of SMM activity pattern over time for one person is overcome later by Rad et el. [64]. Rad et el. [64] selected features used in detecting SMM behaviors based on deep convolutional neural networks .

Applying deep learning techniques for detecting self-stimulatory behaviors has proven its dominance over traditional techniques. Although that the behavioral patterns vary from subject to another in terms of speed frequency, Rad et al. [64] have proposed neural network model trained on large observations samples to overcome the inter-subject variability. However, still Rad was unable to do so. Therefore T. Gadi et al. [65] proposed a framework that rely on deep learning and requires only few labelled data per person to overcome inter-subject variability. T. Gadi et al. [65] proposed solution that apply deep learning techniques on generic to detect SMM. They used two Convolutional Neural Network models in time and frequency domain with appropriate configurations to SMM signals. T. Gadi et al. have also applied cross-domain transfer learning to an alternative of in-domain transfer learning that enhanced the performance of correctly identify SMM patterns for any new subject by using few labelled data. Although that this research was a state of the art at that point, the proposed system is generic as it applies deep learning methodologies to generalize the work and beat inter-subject variability. However, the main limitation is relying on wearable sensors to collect data. Wearable sensors increase feeling of anxiety in autistic children.

Also, Lamyaa Sadouk et al. [65], have overcome inter-subject variability by utilizing CNN models. Lamyaa Sadouk et al. proposed model applies within-domain and across-domain transfer learning on input features. Lamyaa's framework has several benefits: 1) detect any new pattern of SMM behaviors with few labelled data 2) overcome the need of huge amount of labelled data by applying comprehensive learning techniques which can detect new patterns without the need of similar pattern. Hadi Moradi et al. [66] have proposed a creative idea to automatically detect children suspected to be diagnosed with autism. Hadi's proposed idea had a great impact on children under 6 years old who usually spend long time playing with toys and show repetitive patterns. Hence, Hadi et al developed smart car that record and analyze repetitive and stereotypical movements while the child is playing with it. The toy car is supported with an 3D accelerometer recording subject's activities. The smart car has been tried on 25 autistic youngsters and 25 typical kids as the test and control bunches separately. Support Vector Machine (SVM) is utilized to distinguish between autistic and typical kids. Hadi's framework has achieved 85% correct classification, 93% sensitivity and 76% specificity. The outcomes were balanced between boys and girls showing the conceivable boundless utilization of this framework among all youngsters.

e) Approaches for detecting autism using vision techniques

Researcher and scientists have believed in vision techniques long time ago. The process of computer-vision based disorder detection goes through two development stages: the first is recording and manually labelling data, and the second is using computer vision techniques coupled with machine learning to automatically detect and recognize disorders .

1. Detecting autism using manual vision

The use of home videos to help in detection and recognition as well as assisting clinicals was an aim for long time. In the 90s, a research [67] of some home movies recorded by parents and care givers were analyzed to recognize autistic symptoms in children; the analysis confirmed the usability of this approach. The researchers claimed that the detection of behavioral deviations is easy during the first two years of the child. Another earlier research [13] is conducted to analyze home videos of children before autism detection. Researchers aimed to analyze interactions between mother and child through first two years and found some important signs: low mother-infant interaction, less connection between mother and baby and tend to avoidance. Clinicals who analyzed the recorded videos confirmed some specific signs of autism in recorded infants, which in turn opened the door of early treatment. Another

research [68] stated that recorded videos during first birthday shows that "looking at other person" was one of the most affecting single predicators of detecting autism. Many behaviors like (looking, response to name, stereotypies, and communication disorders) have been recognized by manually analyzing recorded videos. Videos were analyzed and classified by specialists into three groups of disorders by taking random cross-sections of scenes and ensuring that the child is visible in all scenes .

2. Detecting autism using computer vision

As the application of computer vision techniques, in cooperation with machine learning, to automatically detect and recognize disorders, is gaining a special interest, the following section is dedicated for reviewing this research direction.

4. Autism Signs Detection And Recognition Using Computer Vision

4.1. Introduction

Many researchers have contributed for detecting self-stimulatory SMM activities in autistic children using one of three approaches: paper-and-pencil measuring ratios, observation session of child's activities, and video-based computerized frameworks. Unfortunately, all abovementioned approaches have many limitations in terms of time and accuracy. However, video-based approach seems more reliable than others, but it still consumes time and effort from clinicals to review videos several times to accurately detect autistic signs. So, automatic techniques are required to overcome the limitation of other approaches .

Applying computer vision techniques to recognize ASD has many benefits, such as reducing clinical expenses in detecting autism and expanding standard tools to detect autism faster and easier.

J. Hashemi et al. [5] have worked on providing a computer vision-based tool to detect and scale ASD special symptoms like visual following, lack of attention, sharing interest and motor patterns. J. Hashemi relied on the Autism Observation Scale for Infants (AOSI) as a reference of atypical autism properties [69]. J. Hashemi et al. aimed to assist in detecting autism biomarkers from actual clinical recorded videos without need of human intervention. J. Hashemi et al. [71] also have proposed a cheap and effective computer vision approach of detecting autism signs during clinical assessment. J. Hashemi et al. focused mainly on two autistic signs belonging to the measures of Autism Observation Scale for Infants (AOSI) [72], which are the lack of attention and Visual following. They compared their results with the domain experts and non-experts and showed that their proposed tool is efficient to detect critical autism signs .

One recent research [4] has suggested using computer vision-based framework to detect early symptoms of autism in young children. Autistic children reactions were recorded, analyzed, and compared to typically developing children's reaction based on tasks assigned to both. The analysis showed variance in body and eye movement response-time and humero-radial directions. B. Noris et al. [8] focused his research on visual behaviors especially eye gaze. B. Noris et al. aimed to calculate the eye gaze direction and understand what child is gazing at during dyadic interaction. Basilio Noris et al. studied several metrics related to gaze, such as the frequency and duration of looking to faces, how significantly the stare reveals the broad field of view and the use of central vs. peripheral vision .

A review of research of eye tracking as an early sign of autism has been done by researchers who claims the applicability of relying on eye tracking for autism detection in infants [13]. Eye tracking in cooperation with machine learning and neural network was a successful method to automatically detect autism .[73]

J. M. Rehg et al. [74] have introduced approaches of analyzing social and communicative video recorded behaviors between children aged 1-2 year and adults by using Rapid-ABC approach [75]. J. M. Rehg et al. worked out to measure the level of engagement which is a red flag of developmental and behavioral disorders. J. M. Rehg et al. have also created the Multimodal Dyadic Behavior (MMDB) dataset that contains more than 160 videos of structured child-adult interaction [3]. Behavioral imaging

can be used to understand various activities. Accordingly, J. M. Rehg et al. have proved that applying behavioral imaging has a great impact on detecting developmental disorders like autism .

Missing one of developmental milestones in children might be a clear indication of developmental disorders. For that reason, Prego et al. [78] focused his research on analyzing recorded milestones video to detect any missing. The features Prego relied on include time spent to do the action and response, as well as the amount of movement adjustment. Prego utilized infra-red camera and used Support Vector Machine (SVM) classifier of children movements.

4.2. Methodologies

Repetitive actions represent one of the most crucial indicators of autism. Among these repetitive behaviors that reveal autism in young children; self- stimulatory or stereotypy behaviors are one of the most important signs of autism. Self- stimulatory are repeated movements of body parts or objects [22]. Stereotypical Motor Movements (SMM) in autism (for example, body shaking, mouthing, and hand flapping [23]) can delays educational and social development. These behaviors can be recognized from recorded videos especially because it follows some pattern. However, it is not straightforward process specifically in uncontrolled environment.

Computer vision based human activity recognition

Human activity recognition is one of the valuable research areas in computer vision field for the reason of their wide scope such as Human Computer Interaction (HCI) [79]. One approach of human activity recognition is by using wearable sensors or smartphones as it records daily activities by using built in accelerometers. However, in many circumstances, wearables and smartphones are not the best approach. Accordingly, Computer vision-based methods can alternate wearable sensors. Vision recording tools have been improved recently in many aspects like recoding depth details and 3D images. J. K. Aggarwal et al. [80] have proposed techniques for human activity recognition using 3D data and indicated its usability in the future.

To detect behaviors of autistic children, computer vision based human activity recognition goes through four main phases. These phases are: Image Segmentation to physical objects, extraction of the features that identify the physical object, body modelling and finally recognizing and classifying activities .

4.2.1. Image Segmentation to Physical Objects

Image segmentation is the fundamental step of human activity recognition which aims to find demanded objects from scene. Usually required objects are in the foreground and some information about the background are known in advance. There are two types of segmentation methods: background construction-based segmentation and foreground extraction-based segmentation. Background subtraction or sometimes called Background construction-based techniques are the most common segmentation methods. Initially, background model which contains only static background scene is created, then any later modification that appears on the scene is considered as moving object. This method is useful for quick-moving objects detection. The main feature of background subtraction method is its low computation cost. However, background subtraction technique is not producing accurate results with moving cameras or if there are no prior information about the background .

In case that camera, as well as objects, are moving, the background information cannot be known in advance which make the phase of segmentation more challenging. Temporal, spatial, or spatio-temporal information are used to extract aimed object from the foreground. In successive frames, aimed object is detected using difference between frames, motion information, or any other feature-based information [81]. Essential characteristics are extracted from the images after segmentation phase and used as features to represent the demanded objects which in turn will be used to classify the objects .

4.2.2. High Level Feature Extraction and Representation

The second phase of human activity recognition is feature extraction and representation. After extracting the objects from the images, they are characterized as a set of features like form, color, or

gesture features. Feature extraction techniques have three families which are global, local, and semantic based features. Global feature extraction method deals with the image as a whole feature, in local feature extraction the processing is concerned with some pixels. Semantic-based features method deals with more complicated human features like posture or simple actions .

Semantic based techniques aim to understand human activities like human brain in which humans understand activities by analyzing human body, posture, and the environment around the activity. Therefore, Semantic based techniques are mostly used for video based HAR. In semantic-based action recognition, activities' information is added in advance based on human understanding. The used information is formed in human body models -descriptors- to help in extracting people from the scene .

The first step in semantic based techniques is to find aimed person in video by using the human body models and then detect the location and some details about the surrounded place. Semantic-based methods have four subcategories; shape-based, appearance-based, pose-based, and motion-based features. Successful human detection and therefore activity recognition rely on accurate human body description. Shape features aim to express the object's shape in term of edges. J. De Winter et al [82] have proved that humans brains differentiate between objects by detecting their edges .

Appearance-based features aim to express objects in terms of colors and surface type rather than edges in shape features. Appearance based methods are usually used for tracking humans in videos in which human body structure is constructed based on consecutive frames of the video. Then, human body model can be tracked. Appearance based methods can give more details from images and produce better results with occlusion compared with shape-based method. This methodology has several benefits: first: extracting the human body model and neglect the background information with no need to apply background subtraction methods. Second: easiness of processing needed. Finally: this method can consider the surrounding atmosphere in recognizing the activity. However, these models are delicate to attire and brightening changes .

Pose-based methods rely on detecting and recognizing human poses, then estimating the activity that correspond to detected sequence of poses. However, same actions can be performed in several poses which makes the process of estimating the pose challenging. T. Moeslund et al. [83] have proposed a classifier of pose estimation based on supervised and unsupervised learning. Supervised pose estimation methods are divided into two classes based on the type of used human body model. Direct model utilizes geometry and kinematics sciences to recalculate representation of the body pose, while indirect model is 3D constructed model of the human body for poses guesstimate. In unsupervised, so called model free methods, no prior body model is used. Model free approach uses transforming sequence of 2D images into 3D pose. Pose-based techniques is making the activity learning process easier because they extract high level features [84]. Adding to that, pose-based techniques are robust against changing the angle of recording. However, recognizing different poses in real situation is the main restriction.

Motion is the core variance between videos and static images. Motion features add a new dimension of information to the descriptor, in which it is used to define the type motion of objects if exist. Motion features are used to differentiate between moving and static objects. Temporal Difference and optical flow are the most known techniques used to extract motion features. Temporal difference method is one of the most popular segmentation techniques for moving camera. Detecting objects in this method is done by first detecting the motion of camera, because both objects and cameras are moving, then subtracting two consecutive frames of the recorded video at time t and t-1, to detect the moving objects, using.(1)

$$|f_t (x,y)-f_(t-1) (x,y)|>T \ \& \ |f_(t+1) (x,y)-f_t (x,y)|>T \quad (1)$$

As t, t-1 and t+1 are current, previous, and next time instant, while f (x, y) represents the features in frame.

Kim et al. [85] proposed a temporal difference method to recognize the movements of the camera using edges features, which helped in turn to position the images correctly based on the noticed movement type and value. Calculating image motion is done by considering each pixel as a motion pixel if the change with next frame exceeds certain threshold. Temporal difference methods are efficient according to its processing simplicity. However, camera motion should be calculated in advance .

The second technique is optical flow, which can be described as a vector of points that describe the motion between frames. Optical flow is used to notice the motion between two successive frames .

Since each type of features extractors target different information than others, it can be said that using different types of features together could produce stronger descriptor. To sum up, although that each type of features provides very useful information, but shape features were preferred to many researchers in the publications of last decade. Shape features have proved its efficiency in object detection since it is the basic way that humans discriminate between different objects.

4.2.2.1. Body modelling

According to [83][86], there are three ways for describing human body; model free, indirect, and direct models. Further processing is performed on these models to extract additional efficient form of features. The successful human detection and therefore activity recognition depends on the accurate human body model. The phase of body modelling (human body descriptor building) is placed before activity recognition phase .

a. Model free

In model free, there is no constructed human body model, the human body is tracked by placing some markers on it, these markers are used for tracking the body. Therefore, this class of models requires intensive training process to enable automatic accurate detection of the markers. The used markers can be in many forms like a shape that surround the human body, spots on the main joints in the body or even stick-man graph on the skeleton. Several researchers have used this model to trail human body in different applications. In [87] Morishima et al. have used ellipse to surround the human body region and track it. In [88][89] researchers have invented algorithms for placing spots on the main joints of the limbs, then track the markers. In [90], stickman is used to represent the human skeleton and track it.

b. Indirect model

In indirect model descriptors, models are performing some secondary function, in which models are used to guide in clarifying extracted data. In this class, the used models lack for deep details which in turn make the process of detecting complex body situation very hard.

c. Direct model

In direct model descriptors, human body model is constructed in advance to detect human bodies in the scene. Direct model is updated repeatedly during the detection process.

4.2.2.2. Human body Model construction

Constructing efficient human body descriptor plays a vital role in accurate human body detection and extraction. However, the construction of human body descriptor is not an easy process and requires collecting the effective features from local regions in the images. Several techniques were proposed to handle this point. One approach is Grid-based human body model construction, which aims to dividing the image into a grid. Another technique aims to extract features from striking selected regions, this technique is called Point-based human body model construction. Two techniques are followed to construct human body descriptors, which are global and local construction techniques. Point and Grid-based can be used in either technique. Global construction aims to construct a model for the whole object, while local approach constructs descriptors for parts, then collected to construct the whole object descriptors. Global approach is easier in implementation and it provides promising results. The contribution proposed by Dalal et al. [91][92] was leading in human body detection on INRIA dataset in which they followed the global approach with HOG methods. Although that global approach succeeded

in many applications like pedestrian detection [93][94], it lacks for parts details of the objects. The global approach is unlikely to be used if the details about body parts affect the output of the application like gesture understanding applications. Although parts details specification was a lack in global approach, it is a benefit in part-based approach. Based on part-based construction approach, an idea named "poselet" was proposed to describe human body parts in discriminative way and therefore used to detect human body poses .

4.2.3. Activity Recognition and Classification

The output of the previous two layers is utilized to detect and recognize the activities. After extracting the suitable features that describe human body, classification algorithms do the job by detecting whether the extracted features are human or non-human object. Support Vector Machine is one of the widely used binary classification algorithms that do so by maximizing the margin between two classes. After ensuring that extracted features represent human, activity recognition methods are used to detect and identify the actions. Human activity recognition relies on using proper classification techniques to assign each extracted body or series of them into a specific predefined class of actions and activities. Human activity recognition phase includes two main steps, which are activity recognition and activity pattern discovery. Activity recognition detects human activities precisely according to a model that describe this activity. Activity pattern discovery is more about discovering some obscure patterns specifically from low-level sensor information with no predefined models or suspicions. Although the two concepts are distinct, both help in producing better accuracy of human activity recognition. Furthermore, they are corresponding to one another - the found activity pattern can be utilized to characterize the activities that will be recognized and followed. Human activity recognition systems have several types depending on the people activities in the videos.

There are three main categories of HAR systems, Single person, many people interactions and suspicious behaviors as shown in Figure 2. The first category is considered as the simplest as well as the most important type, because it is used as the fundamental step of recognizing the activities in other types .

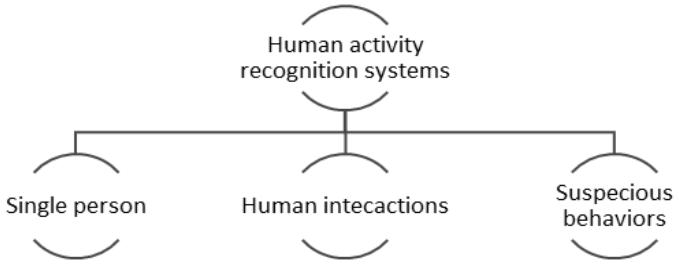


Figure 2 HAR systems categorization based on number of participants

Binary Classification algorithms are used to ensure the accuracy of the detection algorithms, in which detected region is classified into human and non-human body. Then another phase of understanding the situation of detected human body is needed in which multi-class classification algorithms are used. Deep learning and convolutional neural networks have achieved great results in computer vision field especially in object detection [95] and image classification [96]. One of the amazing benefits of using deep neural networks is its ability of automatically choosing object descriptors and to learn complicated features through training phases. Lingfei Mo et al. [97] proposed a computer vision based human activity recognition framework based on deep learning method to identify some physical activities. The activity recognition relies on the skeleton data of the human body which is tracked by the depth sensor

Microsoft Kinect. The Lingfei Mo et al. used CAD-60 dataset which is a dataset of human skeletons. The used model requires no training process which in turn eliminates the data pre-processing and feature extraction steps. The framework recognizes 12 activities with accuracy of 81.8% which indicates the efficiency of convolutional neural network for supervised learning and human physical activity recognition. Gabriel et al. [98] has used deep learning in his research to perform human body segmentation. The solution Gabriel et al proposed relies on Convolutional Neural network in which each pixel is mapped to predefined set of classes of human body parts like head, upper limbs, and lower limps. Also, several successes were achieved in [99], [100] using deep models to detect pedestrians.
The process pf human activity recognition is relying on extracting human body from the input video or image. This step is not as easy as it seems to be done by humans. In human brains we can extract human body in many different situations by just looking at the image, however the mental process is very complicated. Li, Feifei et al. [101] have studied the relationship between how human understand objects and how can we transform this knowledge to computational methods. The various conditions and activities that humans perform make the human body detection step more challenging in which several parameters affect the final appearance of the human body. Some of these parameters are correlated to human body itself, surroundings or even tools used in capturing images of videos.

4.3. Evaluation metrics

One metric of calculating the accuracy of classification algorithms is done by calculating factors of correctly detected human activity recognition with respect to total labelled human activity patterns. Several calculations are used to evaluate the accuracy of classification algorithms. Some are expected to be maximized while others should be minimized. Here we will list some of these calculations.
Metrics used to evaluate the accuracy of classification algorithms, include but not limited to:

- True Positive Rate (TPR): is the ratio between true positive detection and positive labelled samples, it should be maximized.
  $TPR = TP/P = TP/(TP+FN) = 1-FNR$ (2)
- True Negative Rate (TNR): is the ratio between true negative detection and negative labelled samples, it should be maximized.
  $TNR = TN/N = TN/(TN+FP) = 1-FPR$ (3)
- False Positive Rate (FPR): is the ratio between false positive detection and negative labelled samples, it should be minimized.
  $FPR = FP/(TN+FP) = FP/P = 1-TNR$ (4)
- False Negative Rate (FNR): is the ratio between false negative detection and positive labelled samples, it should be minimized.
  $FNR = FN/(TP+FN) = FN/P = 1-TPR$ (5)
- False Discovery Rate (FDR): is the ratio between false positive detection and the summation of true and false positive detection, it should be minimized .
  $FDR = FP/(FP+TP) = 1-PPV$ (6)
- False Omission Rate (FOR): is the ratio between false negative detection and the summation of true and false negative detection, it should be minimized.
  $FOR = FN/(FN+TN) = 1-NPV$ (7)
- Positive Prediction Value (PPV): is the ration between true positive detection and the summation of true and false positive detection, it should be maximized.
  $PPV = TP/(TP+FP) = 1-FDR$ (8)
- Negative Prediction Value (NPV): is the ration between true negative detection and the summation of true and false negative detection, it should be maximized.

NPV= TN/(TN+FN) =1-FOR (9)

- Accuracy (ACC): is the ration between the summation of true positive detection and true negative detection and the total of samples, it should be maximized.
  ACC= (TP+TN)/(P+N) =(TP+TN)/(TP+TN+FP+FN) (10)

5. Human Activity recognition Datasets for Autism detection

To implement human body detection algorithms, several publicly published datasets can be used to evaluate the accuracy of the algorithms. These datasets include, but not limited to:

- **The DE-ENIGMA Database[102]**

The DE-ENIGMA Database of autistic children's interactions is free, multi-modal, and suitable for behavioral and machine learning research, based on the activities of 62 British and 66 Serbian autistic children. This dataset contains 152 hours of video communications among the children and either adults or robots.

- **Autism Spectrum Disorder Detection Dataset[103]**

A new video dataset containing in a set of videos to understand actions performed by autistics and typically developing (TD) kids. Subjects were separated into two group and perform specific tasks. The tasks were to hold a bottle, place it, pouring, passing it to pour and pass it to some place. The recorded videos are processed to classify the child to ASD child or typically developing-TD child based on recent psychological studies and neuroscience.

- **Autism screening data for toddlers | Kaggle[104]**

A recent dataset of ASD screening is published by Dr Fadi Fayez Thabtah. The dataset contains characteristics of 10 behavioral features in addition to other features that have been found effective in detecting autism. The dataset contains of 1054 records with 18 attributes per each record. This dataset can be utilized to evaluate machine learning based Autism behavior understanding frameworks

- **Self-Stimulatory Behaviors in the Wild for Autism Diagnosis Dataset**[22]

A new accessible dataset (SSBD) of self-stimulatory behaviors performed by children to be used to detect autism. The dataset includes 75 recordings with mean time of 90 seconds per video, assembled under three classifications: arm flapping, head banging, and spinning. The videos are recorded in uncontrolled environments which makes them very challenging for automatic detection and recognition .

- **Autistic Spectrum Disorder Screening Data for Children Dataset[105]**

A new autism screening dataset containing 20 features to be utilized for detecting autism red flags and developing the process of ASD detection and recognition. Ten behavioral features (AQ-10-Child) and other 10 attributes were recorded in this dataset. Behavioral sciences use these features to detect and classify autistic cases .

- **National Database for Autism Research (NDAR)[106]**

National Database for Autism Research (NDAR) is a public sharable, accessible domain for biological and behavioral autism data with several different forms and data types. NDAR aims to speed up the progress of research in autism research.

- **A Dataset of Eye Movements for the Children with ASD[107]**

A dataset of Eye Movements for the children with ASD is a dataset that track the eye movements of autistics and typically developed children. The dataset consists of three hundred images of eye tracking which can be used by researchers to analyze and detect visual symptoms of autistic children.

- **ASD Video Glossary [108]**

ASD Video Glossary is web-based reference that spreads the awareness of autism in different ways like defining autism and its symptoms, importance of early diagnosis, actions to be performed after diagnosis and other resources that supports parents, caregivers and autistic people .

- **The Multimodal Dyadic Behavior (MMDB) dataset**[109]

A publicly accessible dataset that can be utilized for further research in the recognition of developmental and behavioral disorders. The dataset consists of 160 video of duration 3-5 minutes for children aged from 15 to 30 months. The dataset concentrates on social attention, non-verbal communication and forth and back interaction. The MMDB dataset can help in detection social and communicational defects which are main symptoms of autistic children.

6. Challenges and Research Objectives

- Building wide dataset facilitating the detection and evaluation techniques. As well as constructing dense dataset covering various indicators variations among autistics .
- Providing an automated assist to the specialist for the process of (early) detection of autistic people by only using videos.
- Design of an objective quantification of the atypical behaviors on which the diagnosis of autism is currently based, through the video analysis of the behaviors of autistic kids and the comparison with the behaviors of children without autism.
- Although that several deep convolutional neural networks-based methods have enhanced action classification outcomes, but there is still a demand to accomplish accurate action localization in videos recorded in uncontrolled environment.

7. Proposed Solution

Based on deep research in this point, we propose a computer vision based neural network model for self-stimulatory behaviors detection. Our proposed solution can be used with publicly available datasets or we can construct newly dataset that contains the aimed behaviors. The proposed model goes through four main stages, which are dataset pre-processing, features extractions, behavior description, and behavior classification as shown in Figure 3 .

Figure 3: Repetitive Motor Behavior Detection Model Block Diagram

First, set of pre-processing operations are required to be performed on the used videos to prepare the data for the latter stages. One of the most important pre-processing operations is noise removal that aims to eliminate distracting information in the videos. Second, aimed features should be extracted from video frames. We propose to use efficient human body pose estimation and skeleton representation neural network models named OpenPose [110]. OpenPose is an open-source neural network model that extract human body features and represents them in form of joints with X&Y coordinates as shown in Figure 4. Third, once the human body joins are extracted, behavioral patterns can be constructed by tracking the aimed features. For example, in case of detecting arm flapping, the arms joins are the aimed joints to be tracked. Finally, once the behavioral patterns are collected from the videos of the dataset, classification neural network model will be used to distinguish between typical and atypical behaviors. The proposed model can be customized to recognize various types of human activities.

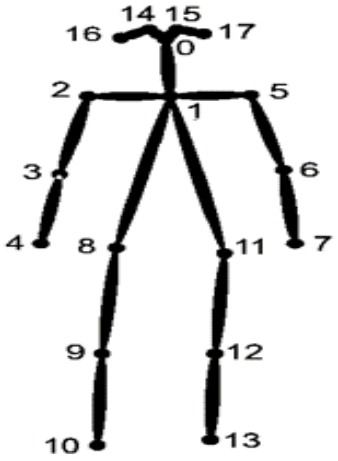


Figure 4 : OpenPose Extracted human body joins

8. Conclusions

Autism spectrum disorder is a complicated mental syndrome manifest itself in children with various severity levels. Although that autism experts are aware of autism symptoms, but the diagnosis process does not always be done easily. Diagnosing autism may take a while, adding to that some cases require many clinical appointments. Not everyone has easy access to these clinics, and people may wait months for an appointment.

As the cause of autism is still unknown, researchers expected that there are some defections in genetics that cause autism. However, genetic based autism diagnosis is still mysterious. For that reason, clinicals usually diagnose autism by monitoring child's behaviors, which is time and effort consuming. Despite the promising results achieved by brain imaging techniques like FMRI or electroencephalography-based protocols, they require complicated provisions that increase the feeling of anxiety in autistic children. Also, brain imaging and EEG-based diagnosis techniques are very expensive ways of diagnosing autism.

Extensive research in the different autism detection approaches and techniques, indicated that although sensor based human activity recognition proved its efficiency, it urges nervous reactions in autistics leading to more severe repetitive motor behaviors that may turn into self-injury .

As the awareness of autism has increased in the last decade, several families, caregivers and medical centers recorded several symptoms of autism that can be used for research directions. Many families and care givers were recording screening all up-normal behaviors of their children. Computer vision-based diagnosis of autism symptoms will help in providing appropriate treatment earlier.

To design computer vision-based system that diagnose autism symptoms, it is expected to go through the main stages of computer vision based human activity recognition. Vision based autism signs detection framework will go through three main stages which are feature extraction, behavior description, and behavioral pattern classification. The first step aims to find the human subject within the frames of the video. The second step aims to describe each behavior as a sequence of actions to produce behavioral patterns. The last step aims to find the parameters that differentiate between different behavioral patterns and to classify any new behavior into one of these classes. Machine learning techniques and neural networks are expected to effectively differentiate between several behavioral patterns .

Using computer vision techniques to recognize ASD has many benefits such as reducing clinical expenses in detecting autism and expanding standard tools to detect autism faster and easier. However,

proposed computer vision programs are not supposed to suppress medical professionals' work. Indeed, computer vision is an operative way for studying the behavior of a large population autistic children as :
1.It allows to capture autism behavioral patterns in a non-intrusive and continuous way over time.
2 .It has minimal expenses and provides objective and quantified evidence of the impact of the different diagnosis methods currently used .
3. It also enables evaluation of the outcomes of (non)-pharmacological treatments.

References

1. C. Lord, "Follow- Up of Two- Year- Old's Referred for Possible Autism," J. Child Psychol. Psychiatry, vol. 36, no. 8, pp. 1365–1382, 1995.
2. G. Lösche, "Sensorimotor and Action Development in Autistic Children from Infancy to Early Childhood," J. Child Psychol. Psychiatry, vol. 31, no. 5, pp. 749–761, 1990.
3. J. M. Rehg, "Behavior Imaging: Using Computer Vision to Study Autism," in MVA2011 IAPR Conference on Machine Vision Applications, 2011.
4. N. A. Khan, M. A. Sawand, M. Qadeer, A. Owais, S. Junaid, and P. Shahnawaz, "Autism Detection using Computer Vision," Int. J. Computer. Sci. Netw. Secur., vol. 17, no. 4, pp. 256–262, 2017.
5. J. Hashemi et al., "A computer vision approach for the assessment of autism-related behavioral markers," 2012 IEEE Int. Conf. Dev. Learn. Epigenetic Robot. ICDL 2012, pp. 1–33, 2012.
6. E. Duchan and D. R. Patel, "Epidemiology of autism spectrum disorders," Pediatric Clinics of North America. 2012.
7. W. L. Stone and K. L. Hogan, "A structured parent interview for identifying young children with autism," J. Autism Dev. Disord., vol. 23, no. 4, pp. 639–652, 1993.
8. B. Noris, "Machine Vision-Based Analysis of Gaze and Visual Context: An Application to Visual Behavior of Children with Autism Spectrum Disorders," Université de Lausanne, 2011.
9. L. Zwaigenbaum, S. Bryson, T. Rogers, W. Roberts, J. Brian, and P. Szatmari, "Behavioral manifestations of autism in the first year of life," Int. J. Dev. Neurosci., vol. 23, no. 2-3 SPEC. ISS., pp. 143–152, 2005.
10. G. Esposito, P. Venuti, F. Apicella, and F. Muratori, "Analysis of unsupported gait in toddlers with autism," Brain Dev., vol. 33, no. 5, pp. 367–373, 2011.
11. A. Klin, W. Jones, R. Schultz, F. Volkmar, and D. Cohen, "Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism," Arch. Gen. Psychiatry, vol. 59, no. 9, pp. 809–816, 2002.
12. G. Golarai, K. Grill-Spector, and A. L. Reiss, "Autism and the development of face processing," Clin. Neurosci. Res., vol. 6, no. 3–4, pp. 145–160, 2006.
13. T. Falck-Ytter, S. Bölte, and G. Gredebäck, "Eye tracking in early autism research," J. Neurodev. Disord., vol. 5, no. 1, p. 28, 2013.
14. I. Noens, I. van Berckelaer-Onnes, R. Verpoorten, and G. van Duijn, "The ComFor: An instrument for the indication of augmentative communication in people with autism and intellectual disability," J. Intellect. Disabil. Res., vol. 50, no. 9, pp. 621–632, 2006.
15. P. T. Shattuck et al. "Timing of identification among children with an autism spectrum disorder: Findings from a population-based surveillance study," J. Am. Acad. Child Adolesc. Psychiatry, vol. 48, no. 5, pp. 474–483, 2009.
16. S. Baron-Cohen et al., "Psychological markers in the detection of autism in infancy in a large population," Br. J. Psychiatry, vol. 168, no. FEB., pp. 158–163, 1996.

17. G. Dawson, D. Hill, A. Spencer, L. Galpert, and L. Watson, "Affective exchanges between young autistic children and their mothers," J. Abnorm. Child Psychol., vol. 18, no. 3, pp. 335–345, 1990.
18. P. Mundy, T. Sherman, and J. Ungerer, "SOCIAL INTERACTIONS OF AUTISTIC, MENTALLY RETARDED AND NORMAL CHILDREN AND THEIR CAREGIVERS," J. Child Psychol. Psychiatry, vol. 27, no. 5, pp. 647–656, 1986.
19. W. L. Stone, K. L. Lemanek, P. T. Fishel, M. C. Fernandez, and W. A. Altemeier, "Play and Imitation Skills in the Diagnosis of Autism in Young Children," Pediatrics, vol. 86, no. 2, pp. 267–273, 1990.
20. K. S. L. Lam and M. G. Aman, "The repetitive behavior scale-revised: Independent validation in individuals with autism spectrum disorders," J. Autism Dev. Disorder., vol. 37, no. 5, pp. 855–866, 2007.
21. R. Landry and S. E. Bryson, "Impaired disengagement of attention in young children with autism," J. Child Psychol. Psychiatry Allied Discip., vol. 45, no. 6, pp. 1115–1122, 2004.
22. S. S. Rajagopalan, A. Dhall, and R. Goecke, "Self-stimulatory behaviors in the wild for autism diagnosis," Proc. IEEE Int. Conf. Computer. Vis., pp. 755–761, 2013.
23. N. M. Rad, A. Bizzego, S. M. Kia, G. Jurman, P. Venuti, and C. Furlanello, "Convolutional Neural Network for Stereotypical Motor Movement Detection in Autism," in 5th NIPS Workshop on Machine Learning and Interpretation in Neuroimaging, 2015, pp. 15–19.
24. A. Bailey, W. Phillips, and M. Rutter, "Autism: Towards an integration of clinical, genetic, neuropsychological, and neurobiological perspectives," Journal of Child Psychology and Psychiatry and Allied Disciplines, vol. 37, no. 1. pp. 89–126, 1996.
25. C. Gillberg et al., "Autism under age 3 years: a clinical study of 28 cases referred for autistic symptoms in infancy.," J. Child Psychol. Psychiatry., vol. 31, no. 6, pp. 921–34, 1990.
26. L. R. Watson and L. M. Marcus, "Diagnosis and Assessment of Preschool Children," in Diagnosis and Assessment in Autism, E. Schopler and G. B. Mesibov, Eds. Boston, MA: Springer US, 1988, pp. 271–301.
27. T. Falck-Ytter and C. von Hofsten, "How special is social looking in ASD. A review," Prog. Brain Res., vol. 189, pp. 209–222, 2011.
28. E. Thelen, "Kicking, rocking, and waving: Contextual analysis of rhythmical stereotypies in normal human infants," Anim. Behav., vol. 29, no. 1, pp. 3–11, 1981.
29. C. Lord et al., "The Autism Diagnostic Observation Schedule-Generic: A standard measure of social and communication deficits associated with the spectrum of autism," J. Autism Dev. Disorder., vol. 30, no. 3, pp. 205–223, 2000.
30. C. Lord, M. Rutter, and A. Le Couteur, "Autism Diagnostic Interview-Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders," J. Autism Dev. Disorder., vol. 24, no. 5, pp. 659–685, 1994.
31. S. E. Bryson, L. Zwaigenbaum, C. McDermott, V. Rombough, and J. Brian, "The autism observation scale for infants: Scale development and reliability data," J. Autism Dev. Disorder., vol. 38, no. 4, pp. 731–738, 2008.
32. S. Ehlers, C. Gillberg, and L. Wing, "A Screening Questionnaire for Asperger Syndrome and Other High-Functioning Autism Spectrum Disorders in School Age Children," J. Autism Dev. Disorder., 1999.
33. J. Williams et al., "The CAST (Childhood Asperger Syndrome Test): Test accuracy," Autism, 2005.
34. D.-5 A. P. Association, "Diagnostic and statistical manual of mental disorders," Arlingt. Am. …, pp. 1–37, 2013.
35. A. A. Baumeister and R. Forehand, "Stereotyped Acts," Int. Rev. Res. Ment. Retard., 1973.

36. G. T. Baranek, “Autism during infancy: A retrospective video analysis of sensory-motor and social behaviors at 9-12 months of age,” J. Autism Dev. Disord., vol. 29, no. 3, pp. 213–224, 1999.
37. L. Wing, “THE HANDICAPS OF AUTISTIC CHILDREN—A COMPARATIVE STUDY,” J. Child Psychol. Psychiatry, vol. 10, no. 1, pp. 1–40, 1969.
38. Y. Shen et al., “Clinical Genetic Testing for Patients with Autism Spectrum Disorders,” Pediatrics, 2010.
39. K. Wang et al., “Common genetic variants on 5p14.1 associate with autism spectrum disorders,” Nature, 2009.
40. J. A. S. Vorstman, W. G. Staal, E. Van Daalen, H. Van Engeland, P. F. R. Hochstenbach, and L. Franke, “Identification of novel autism candidate regions through analysis of reported cytogenetic abnormalities associated with autism,” Molecular Psychiatry. 2006.
41. J. Veenstra-VanderWeele, S. L. Christian, and E. H. Cook, Jr., “AUTISM AS A PARADIGMATIC COMPLEX GENETIC DISORDER,” Annual. Rev. Genomics Hum. Genet., 2004.
42. N. Momeni et al., “A novel blood-based biomarker for detection of autism spectrum disorders,” Transl. Psychiatry, 2012.
43. G. Pusiol, A. Esteva, S. S. Hall, M. Frank, A. Milstein, and F. F. Li, “Vision-based classification of developmental disorders using eye-movements,” in International Conference on Medical Image Computing and Computer-Assisted Intervention, Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016, 2016, vol. 9901 LNCS, pp. 317–325.
44. P. J. Hagerman, “The fragile X prevalence paradox,” Journal of Medical Genetics, vol. 45, no. 8. pp. 498–499, 2008.
45. A. Sheikhani, H. Behnam, M. R. Mohammadi, M. Noroozian, and M. Mohamamadi, “Detection of abnormalities for diagnosing of children with autism disorders using of quantitative electroencephalography analysis,” J. Med. Syst., 2012.
46. M. I. Kamel, K. Abd, and F. Linea, “EEG based Autism Diagnosis Using Regularized Fisher Linear Discriminant Analysis,” Image, Graph. Signa Process., vol. 3, pp. 35–41, 2012.
47. W. Bosl, A. Tierney, H. Tager-Flusberg, and C. Nelson, “EEG complexity as a biomarker for autism spectrum disorder risk,” BMC Med., 2011.
48. G. Chanel, S. Pichon, L. Conty, S. Berthoz, C. Chevallier, and J. Grèzes, “Classification of autistic individuals and controls using cross-task characterization of fMRI activity,” NeuroImage Clin., 2016.
49. N. C. Dvornek, D. Yang, P. Ventola, and J. S. Duncan, “Learning generalizable recurrent neural networks from small task-fMRI datasets,” in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2018.
50. F. Albinali, M. S. Goodwin, and S. S. Intille, “Recognizing Stereotypical Motor Movements in the Laboratory and Classroom: A Case Study with Children on the Autism Spectrum,” Proc. Int. Conf. Ubiquitous Comput., pp. 71–80, 2009.
51. M. S. Goodwin, S. S. Intille, F. Albinali, and W. F. Velicer, “Automated Detection of Stereotypical Motor Movements.,” J. Autism Dev. Disord., vol. 41, no. 6, pp. 770–782, 2011.
52. T. Westeyn, K. Vadas, X. Bian, T. Starner, and G. D. Abowd, “Recognizing mimicked autistic self-stimulatory behaviors using HMMs,” in Ninth IEEE International Symposium on Wearable Computers (ISWC’05), 2005, vol. 2005, pp. 164–167.
53. M. Fan, A. H. Tewfik, Y. Kim, and R. Menard, “Optimal sensor location for body sensor network to detect self-stimulatory behaviors of children with autism spectrum disorder,” in Proc. Int. Conf. Engineering in Medicine and Biology, 2009.

54. C.-H. Min and A. H. Tewfik, "Automatic characterization and detection of behavioral patterns using linear predictive coding of accelerometer sensor data.," Conf. Proc. Annual. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annual. Conf., 2010.
55. C.-H. M. C.-H. Min and a. H. Tewfik, "Novel pattern detection in children with Autism Spectrum Disorder using Iterative Subspace Identification," Acoust. Speech Signal Process. (ICASSP), 2010 IEEE Int. Conf., 2010.
56. C.-H. Min and A. H. Tewfik, "Semi-supervised event detection using higher order statistics for multidimensional time series accelerometer data.," Conf. Proc. ... Annual. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annual. Conf., 2011.
57. F. Albinali, M. S. Goodwin, and S. S. Intille, "Recognizing stereotypical motor movements in the laboratory and classroom," in Proceedings of the 11th international conference on Ubiquitous computing - Ubicomp '09, 2009.
58. F. Albinali, M. S. Goodwin, and S. Intille, "Detecting stereotypical motor movements in the classroom using accelerometry and pattern recognition algorithms," Pervasive Mob. Comput., 2012.
59. M. S. Goodwin, "47.3 EMERGING TECHNOLOGIES FOR MULTIMODAL ASSESSMENT OF AUTISM SPECTRUM DISORDER IN LABORATORY AND NATURALISTIC SETTINGS: UTILITY FOR BIOBEHAVIORAL PHENOTYPING," J. Am. Acad. Child Adolesc. Psychiatry, 2016.
60. R. Fusaroli, I. Konvalinka, and S. Wallot, "Analyzing social interactions: The promises and challenges of using cross recurrence quantification analysis," in Springer Proceedings in Mathematics and Statistics, 2014.
61. V. Romero, P. Fitzpatrick, R. C. Schmidt, and M. J. Richardson, "Using Cross-Recurrence Quantification Analysis to Understand Social Motor Coordination in Children with Autism Spectrum Disorder," in Recurrence Plots and Their Quantifications: Expanding Horizons, 2016.
62. S. Bhat, U. R. Acharya, H. Adeli, G. M. Bairy, and A. Adeli, "Automated diagnosis of autism: In search of a mathematical marker," Rev. Neurosci., 2014.
63. U. Großekathöfer et al., "Automated Detection of Stereotypical Motor Movements in Autism Spectrum Disorder Using Recurrence Quantification Analysis," Front. Neuroinform., vol. 11, no. February 2017.
64. N. Mohammadian Rad et al., "Deep learning for automatic stereotypical motor movement detection using wearable sensors in autism spectrum disorders," Signal Processing, 2018.
65. T. Gadi and E. H. Essoufi, "A Novel Deep Learning Approach for Recognizing Stereotypical Motor Movements within and," Comput. Intell. Neurosci., vol. 2018, 2018.
66. H. Moradi, S. E. Amiri, R. Ghanavi, and B. N. Aarabi, "Ubiquitous Computing and Ambient Intelligence," vol. 10586, pp. 817–827, 2017.
67. J. L. Adrien et al., "Autism and family home movies-Preliminary findings," J. Autism Dev. Disorder., vol. 21, no. 1, pp. 43–49, 1991.
68. J. Osterling and G. Dawson, "Early recognition of children with autism: A study of first birthday home videotapes," J. Autism Dev. Disorder., vol. 24, no. 3, pp. 247–257, 1994.
69. D. Wen, C. Fang, X. Ding, and T. Zhang, "Development of recognition engine for baby faces," in Proceedings - International Conference on Pattern Recognition, 2010, pp. 3408–3411.
70. O. Lanz, "Sampling techniques for audio–visual tracking and head pose estimation," in Multimodal Signal Processing: Human Interactions in Meetings, vol. 9781107022, 2012, pp. 84–102.
71. J. Hashemi et al., "Computer Vision Tools for Low-Cost and Noninvasive Measurement of Autism-Related Behaviors in Infants," Autism Res. Treat., vol. 2014, pp. 1–12, 2014.

72. S. E. Bryson and L. Zwaigenbaum, “Autism Observation Scale for Infants,” in Comprehensive Guide to Autism, V. B. Patel, V. R. Preedy, and C. R. Martin, Eds. New York, US: Springer, New York, NY, 2014, pp. 299–310.
73. R. C. B, F. Cilia, G. Dequen, J. Bosche, J. Guerin, and L. Vandromme, “Automatic Autism Spectrum Disorder Detection Thanks to Eye-Tracking and Neural Network-Based Approach Romuald,” in Internet of Things (IoT) Technologies for HealthCare, 2018, vol. 225, pp. 75–81.
74. J. M. Rehg et al., “Decoding Children’s Social Behavior,” in 2013 IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 3414–3421.
75. J. Mathys, “Beyond Parental Report: Findings from the Rapid-ABC, A New 4-Minute Interactive Autism,” 2013.
76. P. Kohli, J. Rihan, M. Bray, and P. H. S. Torr, “Simultaneous segmentation and pose estimation of humans using dynamic graph cuts,” Int. J. Comput. Vis., vol. 79, no. 3, pp. 285–298, 2008.
77. M. Eichner, M. Marin-Jimenez, A. Zisserman, and V. Ferrari, “2D articulated human pose estimation and retrieval in (almost) unconstrained still images,” Int. J. Comput. Vis., vol. 99, no. 2, pp. 190–214, 2012.
78. P. Perego, S. Forti, A. Crippa, A. Valli, and G. Reni, “Reach and throw movement analysis with support vector machines in early diagnosis of autism,” Eng. Med. Biol. Soc. 2009. EMBC 2009. Annual. Int. Conf. IEEE, 2009.
79. Y. Du, F. Chen, and W. Xu, “Human interaction representation and recognition through motion decomposition,” IEEE Signal Process. Lett., 2007.
80. J. K. Aggarwal and L. Xia, “Human activity recognition from 3D data: A review,” Pattern Recognition. Lett., vol. 48, pp. 70–80, 2014.
81. A. Bux, P. Angelov, and Z. Habib, “Vision Based Human Activity Recognition: A Review,” in Advances in Computational Intelligence Systems, 2016, vol. 513, pp. 341–371.
82. J. De Winter and J. Wagemans, “Contour-based object identification and segmentation: Stimuli, norms and data, and software tools,” Behav. Res. Methods, Instruments, Comput., vol. 36, no. 4, pp. 604–624, 2004.
83. T. B. Moeslund, A. Hilton, and V. Kr??ger, “A survey of advances in vision-based human motion capture and analysis,” Comput. Vis. Image Underst., vol. 104, no. 2-3 SPEC. ISS., pp. 90–126, 2006.
84. A. Yao, J. Gall, G. Fanelli, and L. Van Gool, “Does Human Action Recognition Benefit from Pose Estimation?” Procedings Br. Mach. Vis. Conf. 2011, 2011.
85. Kye Kyung Kim, Soo Hyun Cho, Hae Jin Kim, and Jae Yeon Lee, “Detecting and tracking moving object using an active camera,” 2008, pp. 817–820.
86. T. B. Moeslund and E. Granum, “A survey of computer vision-based human motion capture,” Comput. Vis. Image Underst., 2001.
87. A. Nakazawa, H. Kato, and S. Inokuchi, “Human tracking using distributed vision systems,” Pattern Recognition, 1998. Proceedings. Fourteenth Int. Conf., 1998.
88. C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland, “P finder: real-time tracking of the human body,” IEEE Trans. Pattern Anal. Mach. Intell., 1997.
89. A. Dargazany and M. Nicolescu, “Human body parts tracking using torso tracking: Applications to activity recognition,” in Proceedings of the 9th International Conference on Information Technology, ITNG 2012, 2012.
90. S. Iwasawa, K. Ebihara, J. Ohya, and S. Morishima, “Real-time human posture estimation using monocular thermal images,” in Proceedings - 3rd IEEE International Conference on Automatic Face and Gesture Recognition, FG 1998, 1998.

91. N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005, 2005.
92. N. Dalal, "Finding People in Images and Videos," 2006.
93. R. Benenson, M. Mathias, T. Tuytelaars, and L. Van Gool, "Seeking the strongest rigid detector," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2013.
94. S. Zhang, C. Bauckhage, and A. B. Cremers, "Informed haar-like features improve pedestrian detection," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2014.
95. R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2014.
96. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," Commun. ACM, vol. 60, no. 6, pp. 1097–1105, 2012.
97. L. Mo, F. Li, Y. Zhu, and A. Huang, "Human physical activity recognition based on computer vision with deep learning model," in Conference Record - IEEE Instrumentation and Measurement Technology Conference, 2016, vol. 2016-July.
98. G. L. Oliveira, A. Valada, C. Bollen, W. Burgard, and T. Brox, "Deep learning for human part discovery in images," Proc. - IEEE Int. Conf. Robot. Autom., vol. 2016-June, pp. 1634–1641, 2016.
99. P. Luo, Y. Tian, X. Wang, and X. Tang, "Switchable deep network for pedestrian detection," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2014.
100. X. Zeng, W. Ouyang, M. Wang, and X. Wang, "Deep learning of scene-specific classifier for pedestrian detection," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2014.
101. F. Li, "Visual Recognition: Computational Models and Human Psychophysics," Itinerario, 2003.
102. "DE-ENIGMA Database." [Online]. Available: http://de-enigma.eu/resources/the-de-enigma-database./
103. "Autism Spectrum Disorder Detection Dataset." [Online]. Available: https://pavis.iit.it/datasets/autism-spectrum-disorder-detection-dataset.
104. Fadi Fayez Thabtah, "Autism screening data for toddlers | Kaggle." [Online]. Available: https://www.kaggle.com/fabdelja/autism-screening-for-toddlers/version/1.
105. "Autistic Spectrum Disorder Screening Data for Children Data Set." [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Children .++
106. "National Database for Autism Research (NDAR)." [Online]. Available: https://healthdata.gov/dataset/national-database-autism-research-ndar.
107. "A Dataset of Eye Movements for the Children with Autism Spectrum Disorder." [Online]. Available: https://zenodo.org/record/2647418#.XT8jDugvPDc.
108. "ASD Video Glossary." [Online]. Available: https://www.autismspeaks.org/what-autism. [Accessed: 05-Jun-2020].
109. "The Multimodal Dyadic Behavior (MMDB) dataset." [Online]. Available: http://www.cbi.gatech.edu/mmdb/.
110. Z. Cao, G. Hidalgo Martinez, T. Simon, S.-E. Wei, and Y. A. Sheikh, "OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields," IEEE Trans. Pattern Anal. Mach. Intell., pp. 1–1, 2019.