SJIF Impact Factor 2022: 8.197| ISI I.F. Value:1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 7 | Issue: 5 | May 2022 - Peer Reviewed Journal

AGGRESSIVE ACTION IDENTIFICATION IN AUTISM SPECTRUM DISORDER USING VIDEO ANALYSIS

Shanmughapriya M[1], Poojashree S[2], Monica G[3], Jayanthi K[4]

[1]*Assistant Professor, Department of Information Technology, Sri Sai Ram Institute of Technology, Chennai, India*

[2]*Final Year B Tech, Department of Information Technology, Sri Sai Ram Institute of Technology, Chennai, India*

[3]*Final Year B Tech, Department of Information Technology, Sri Sai Ram Institute of Technology Chennai, India*

[4]*Final Year B Tech, Department of Information Technology, Sri Sai Ram Institute of Technology Chennai, India*

Article DOI: https://doi.org/10.36713/epra9981
DOI No: 10.36713/epra9981

ABSTRACT

Autism Spectrum Disorder ASD is a clinical condition associated with brain development. It affects how a person perceives the world, Pathological symptoms of ASD varies from mild to severe. Repetitive action or pattern of action is often associated with ASD. One of the main concerns of autistic children is the aggressive behavior that can be directed towards them, which may lead to self injuries. Constant monitoring is required in case of autism with aggressive behavior to prevent self injuries. The proposed work aims to recognize three classes of violent action namely head banging against wall or any object, arm flapping continuously and spinning uncontrollably from video data. The proposed work uses a 3DCNN model with Skeleton Joint features for recognizing the said actions. The accuracy of proposed model is 83.56% in dataset validation. The cross data accuracy is about 65%. The proposed work also aims at analyzing the difficulties in live video analysis and recognition of actions from live stream video. The self-stimulatory behavior dataset – SSBD is used is this work.

KEYWORDS— *3DCNN, Deep Learning, Action Recognition, Autism Spectrum Disorder*

I. INTRODUCTION TO ASD

Autism spectrum disorder is a condition related to brain development that impacts how a person perceives and socializes with others, causing problems in social interaction and communication. The disorder also includes limited and repetitive patterns of behavior. The term "spectrum" in autism spectrum disorder refers to the wide range of symptoms and severity. Autism spectrum disorder includes conditions that were previously considered separate — autism, Asperser's syndrome, childhood disintegrate.

Autism spectrum disorder (ASD) is a developmental disability caused by differences in the brain. Some people with ASD have a known difference, such as a genetic condition. Other causes are not yet known. Scientists believe there are multiple causes of ASD that act together to change the most common ways people develop. We still have much to learn about these causes and how they impact people with ASD. People with ASD may behave, communicate, interact, and learn in ways that are different from most other people. There is often nothing about how they look that sets them apart from other people. The abilities of people with ASD can vary significantly. For example, some people with ASD may have advanced conversation skills whereas others may be nonverbal. Some people with ASD need a lot of help in their daily lives; others can work and live with little to no support.

ASD begins before the age of 3 years and can last throughout a person's life, although symptoms may improve over time. Some children show ASD symptoms within the first 12 months of life. In others, symptoms may not show up until 24 months of age or later. Some children with ASD gain new skills and meet developmental milestones until around 18 to 24 months of age, and then they stop gaining new skills or lose the skills they once had.

SJIF Impact Factor 2022: 8.197| ISI I.F. Value:1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 7 | Issue: 5 | May 2022 - Peer Reviewed Journal

As children with ASD become adolescents and young adults, they may have difficulties developing and maintaining friendships, communicating with peers and adults, or understanding what behaviors are expected in school or on the job. They may come to the attention of healthcare providers because they also have conditions such as anxiety, depression, or attention-deficit/hyperactivity disorder, which occur more often in people with ASD than in people without ASD.

A. Signs and Symptoms

People with ASD often have problems with social communication and interaction, and restricted or repetitive behaviors or interests. People with ASD may also have different ways of learning, moving, or paying attention. It is important to note that some people without ASD might also have some of these symptoms. For people with ASD, these characteristics can make life very challenging.

II. AUTOMATED VIDEO ANALYSIS – AN OVERVIEW

Humans easily recognize and identify actions in video but automating this procedure is challenging. Human action recognition in video is of interest for applications such as automated surveillance, elderly behavior monitoring, human-computer interaction, content-based video retrieval, and video summarization. In monitoring the activities of daily living of elderly, for example, the recognition of atomic actions such as “walking”, “bending”, and ”falling” by itself is essential for activity analysis. With the development of society, the estimated ASD prevalence is dramatically increasing. Basically, the traditional diagnosis of ASD aims at improving observation measures and interviewing questions to elicit specific ASD behaviors. It focuses on four key areas: communication, reciprocal social interaction, imagination/creativity, and stereotyped behaviors/restricted interests. However, most of the ASD diagnoses have to be done by specialists make the situation worse.

A. Tools for Human Action Recognition from Video

Over time, computer vision researchers have shifted their focus from image to video, 2D to 3D, and supervised to unsupervised. One of the trends, video understanding, has become a hot topic. Video human action recognition, a basic task within video understanding, also attracts lots of attention. As shown in the timeline refer figure 1, more and more algorithms on video action recognition are proposed each year.

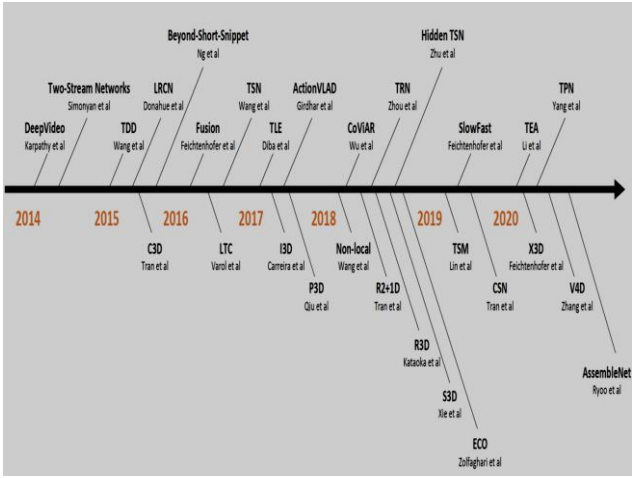

Figure 1 - A chronological overview of recent representative work in video action recognition.

B. Video Analysis and Classification – Need of the hour

Video content analysis or video content analytics (VCA), also known as video analysis or video analytics (VA), is the capability of automatically analyzing video to detect and determine temporal and spatial events. Video management software manufacturers are constantly expanding the range of the video analytics modules available. With the new suspect tracking technology, it is then possible to track all of this subject's movements easily: where they came from, and when, where, and how they moved. Within a particular surveillance system, the indexing technology is able to locate people with similar features who were within the cameras’ viewpoints during or within a specific period of time.

When performing image classification, we:

- Input an image to our CNN
- Obtain the predictions from the CNN
- Choose the label with the largest corresponding probability

Since a video is just a series of frames, a naive video classification method would be to:

SJIF Impact Factor 2022: 8.197| ISI I.F. Value:1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 7 | Issue: 5 | May 2022 - Peer Reviewed Journal

1. Loop over all frames in the video file
2. For each frame, pass the frame through the CNN
3. Classify each frame individually and independently of each other
4. Choose the label with the largest corresponding probability
5. Label the frame and write the output frame to disk

Videos can be understood as a series of individual images; and therefore, many deep learning practitioners would quick to treat video classification as performing image classification a total of N times, where N is the total number of frames in a video. Video classification is more than just simple image classification — with video we can typically make the assumption that subsequent frames in a video are correlated with respect to their semantic contents. If we are able to take advantage of the temporal nature of videos, we can improve our actual video classification results. Neural network architectures such as Long short-term memory (LSTMs) and Recurrent Neural Networks (RNNs) are suited for time series data — two topics that we'll be covering in later tutorials — but in some cases, they may be overkill. They are also resource-hungry and time-consuming when it comes to training over thousands of video files as you can imagine. Instead, for some applications, all you may need is rolling averaging over predictions.

1.Loop over all frames in the video file
2.For each frame, pass the frame through the CNN
3.Obtain the predictions from the CNN
4.Maintain a list of the last K predictions
5.Compute the average of the last K predictions and choose the label with the largest corresponding probability
6.Label the frame and write the output frame to disk

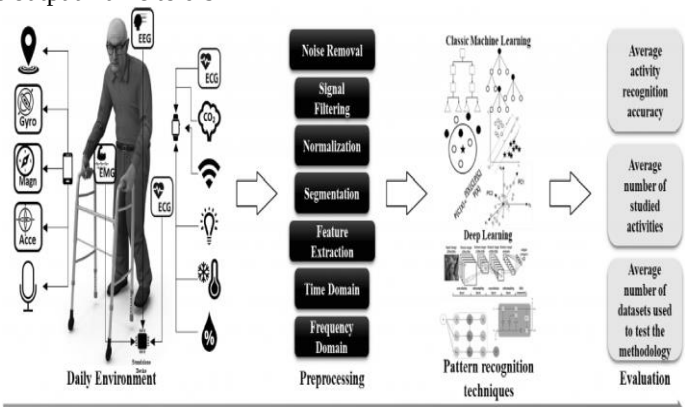

Figure 2 - Steps in Action Recognition from Raw Video

III. SURVEYED LITERATURE

In this section, we first consider related earlier surveysin human action recognition. Looking at the major conferences and journals several earlier surveyshave been published. Aggarwal and Cai reviewedmethods for human motion analysis focusing on three major areas including: motion analysis involving human body parts, tracking a moving human from a single view or multiple cameras and recognizing human activities from image sequences. Moeslund and Granum reviewed papers on human motion capture considering a general structure for systems analyzing human body motion as a hierarchical process with four steps: initialization, tracking, poses estimation and recognition. Wang et al. presented a survey of work on human motion analysis, in which motion analysis was illustrated as a three-level process including human detection (low-level vision), human tracking (intermediatelevel vision), and behavior understanding (high-levelvision). Refer table 1 for surveyed papers.

SJIF Impact Factor 2022: 8.197| ISI I.F. Value:1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 7 | Issue: 5 | May 2022 - Peer Reviewed Journal

Table 1: Surveyed Papers

SL. NO	PAPER NAME AND YEAR	ALGORITHM USED	DATASET USED
1.	Patterns of restricted and repetitive behaviors in autism spectrum disorders: A cross sectional video recording study- May 2021	ADOS SA ADOS RRB	Video recorded by health professional through body cam.
2.	Video based interventions for promoting positive social behavior in children with ASD-2021.	VBI	Video of ASD children behavior at dentistry.
3.	Phonic and motor stereotypes in ASD: Video analysis and neurological Characterization-March 2021	Demographics and clinical Variables.	Observed the ASD children through video and phonic stereotypes are classified.(Similar to MRI scan)
4.	Behavior- Based Risk detection of ASD through child-robot interaction-2020 [1].	CNN (Robots)	Data extracted from video analysis of child-robot interactions.
5.	An explorative study on robotics for supporting children with ASD during clinical procedures-2020.	NAO Robots. (Humanoid)	Videos of ASD children interaction with robots and survey of care givers.
6.	A systematic review of remote tele health assessments for early signs of ASD: Video and mobile applications- May 2020.	Mobile or Web applications.	Mobile home video portal.
7.	Feature replacement methods enable reliable home video analysis for machine learning detection of autism-2020.	Attention deficit Hyperactivity disorder classification algorithm.	ADI-electronic health records.

Moeslund et al. described the work in human capture and analysis based on 280 papers from 2000 to 2006, centered on initialization of human motion, tracking, pose estimation, and recognition. Turaga et al. considered that “actions” are characterized by simple motion patterns typically executed by a single person while “activities” are more complex and involve coordinated actions among a small number of humans and reviewed the major approaches for recognizing human action and activities. Poppe focused on image representation and action classification methods. A similar survey by Weinland et al. also concentrated on approaches for action representation and classification. Popoola and Wang presented a survey focusing on contextual abnormal human behavior detection for surveillance applications. Ke et al. reviewed human activity recognition methods for both static and moving cameras, covering many problems such as feature extraction, representation techniques, activity detection and classification. Aggarwaland Xia presented a survey of human activity recognition based on 3D data, especially on using RGB and depth information acquired by 3D sensors as the Kinect. Guo and Lai gave a survey of existing approaches on still image-based action recognition. Recently, Cheng et al. reviewed approaches on human action recognition using an approach-based taxonomy, in which all methodologies are classified into Two categories: single-layered approaches and hierarchical approaches. In addition, Vrigkas et al. categorized human activity recognition methods into two main categories including “unimodal” and “multimodal”.

SJIF Impact Factor 2022: 8.197| ISI I.F. Value:1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 7 | Issue: 5 | May 2022 - Peer Reviewed Journal

Dataset Name	Color	Depth	Skeleton	Samples	Classes
Hollywood2	√	×	×	1707	12
HMDB51	√	×	×	6766	51
Olympic Sports	√	×	×	783	16
UCF50	√	×	×	6618	50
UCF101	√	×	×	13,320	101
Kinetics	√	×	×	306,245	400
MSR-Action3D	×	√	√	567	20
MSR-Daily Activity	√	√	√	320	16
Northwestern-UCLA	√	√	√	1475	10
UTD-MHAD	√	√	√	861	27
RGBD-HuDaAct	√	√	×	1189	13
NTU RGB+D	√	√	√	56,880	60

Table 2: Bench Mark Dataset for Human Action Recognition

IV. PROPOSED SYSTEM

Studies on autistic children have found some unusual behavior and response by an ASD child. As automated detection aims to automate the whole diagnosis process, so the system needs to focus on those identifying characteristics, e.g., repetitive behavior, atypical walking style, and particular visual saliency. In this section, we have explored the recent literature on activity analysis based ASD detection by arranging them in three main groups. Figure 4 shows the overall process of the activity analysis-based automated detection, followed by the existing approaches in this domain. The activity analysis-based automated detection of ASD comprises three main steps: data collection, training, and classification. The data collection block shows three core approaches; each of them utilizes different sensors to capture different exclusive characteristics of autistic individuals: repetitive behavior, atypical gait pattern, and unusual visual saliency. Next comes the training phase, which utilizes machine learning, or deep learning approaches to learn discriminative features of the data to classify ASD and TD in the final step of the automated detection. The experimental setup for data collection in activity analysis is preferred to be performed in constrained environments. It minimizes the additional features introduced in computational modeling due to different backgrounds or surroundings this makes the model more focused on the specific activity' dynamic features, which is to be analyzed. Consequently, this compels it to be well trained in detecting the desired activity primarily. Further computational development to this basic model with complex data can improve its capabilities to perform well in different environments.

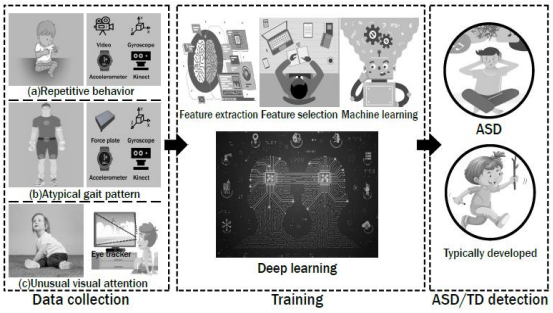

Figure 3: Process of ASD activity analysis

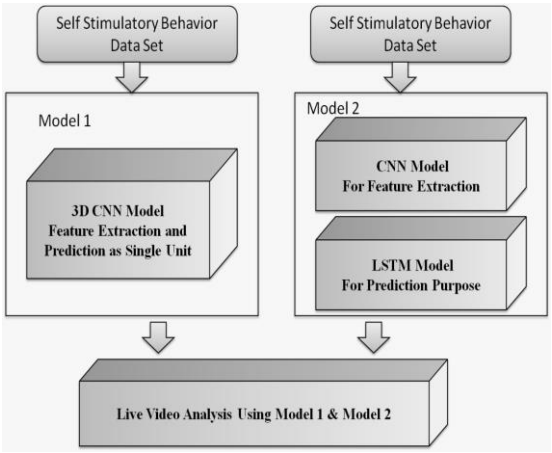

Figure 4. Architecture of the Proposed Work

SJIF Impact Factor 2022: 8.197| ISI I.F. Value:1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 7 | Issue: 5 | May 2022 - Peer Reviewed Journal

Another modality to model human actions are applied the 3D filter to the CNNs. They seem like the standard CNN, but with the spatial-temporal kernel which directly aggregated hierarchical representations of spatial-temporal data. However, the image-based action recognition model is sensitive to a noisy background with complex occlusion or illumination conditions, changing camera view angles. The ability of the classification of this framework was also heavily dependent on the object in the images. For example, when there is a horse in the image, the model usually tends to mistakenly classify the action as horse riding, which means the model focus more on the object instead of the essential action. Moreover, It is expensive to train a 3D CNN, since the 3D filter has many more parameters than the 2D one. A 3D CNN is simply the 3D equivalent: it takes as input a 3D volume or a sequence of 2D frames (e.g. slices in a CT scan), 3D CNNs are a powerful model for learning representations for volumetric data. A convolutional neural network consists of an input layer, hidden layers and an output layer. In any feed-forward neural network, any middle layers are called hidden because their inputs and outputs are masked by the activation function and final convolution. In a convolutional neural network, the hidden layers include layers that perform convolutions. Typically this includes a layer that performs a dot product of the convolution kernel with the layer's input matrix. This product is usually the Frobenius inner product, and its activation function is commonly ReLU. As the convolution kernel slides along the input matrix for the layer, the convolution operation generates a feature map, which in turn contributes to the input of the next layer. This is followed by other layers such as pooling layers, fully connected layers, and normalization layers.

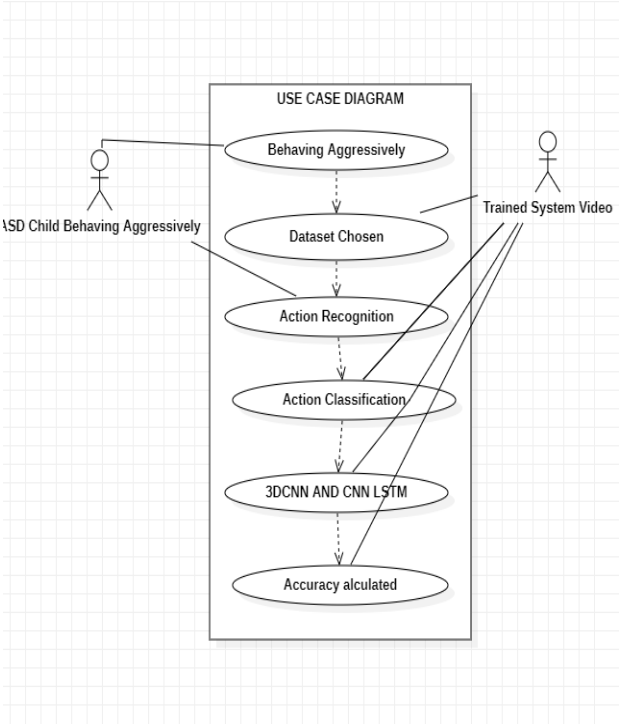

Figure 5. Use Case Diagram for proposed model

A workflow diagram (also known as a workflow) provides a graphic overview of the business process. Using standardized symbols and shapes, the workflow shows step by step how your work is completed from start to finish. It also shows who is responsible for work at what point in the process. Workflows are how you get work done. It's the sequence of tasks from start to finish: The process. Flowcharts are used in designing and documenting simple processes or programs. Like other types of diagrams, they help visualize what is going on and thereby help understand a process, and perhaps also find less-obvious features within the process, like flaws and bottlenecks. There are different types of flowcharts: each type has its own set of boxes and notations. The two most common types of boxes in a flowchart are: A processing step, usually called activity, and denoted as a rectangular box. A decision usually denoted as a diamond. A flowchart is described as "cross-functional" when the chart is divided into different vertical or horizontal parts, to describe the control of different organizational units. A symbol appearing in a particular part is within the control of that organizational unit. A cross-functional flowchart allows the author to correctly locate the responsibility for performing an action or making a decision, and to show the responsibility of each organizational unit for different parts of a single process.

SJIF Impact Factor 2022: 8.197| ISI I.F. Value:1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 7 | Issue: 5 | May 2022 - Peer Reviewed Journal

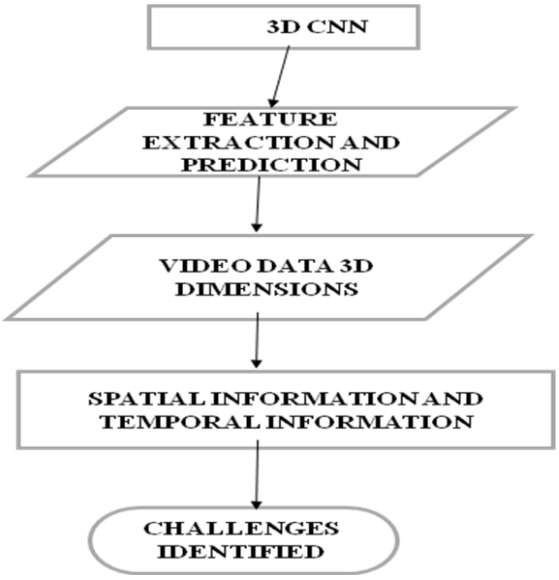

Figure 6: Work Flow of Proposed System

Action recognition task involves the identification of different actions from video clips (a sequence of 2D frames) where the action may or may not be performed throughout the entire duration of the video. This seems like a natural extension of image classification tasks to multiple frames and then aggregating the predictions from each frame. Despite the stratospheric success of deep learning architectures in image classification (ImageNet), progress in architectures for video classification and representation learning has been slower.

Huge Computational Cost A simple convolution 2D net for classifying 101 classes has just ~5M parameters whereas the same architecture when inflated to a 3D structure results in ~33M parameters. It takes 3 to 4 days to train a 3DConvNet on UCF101 and about two months on Sports-1M, which makes extensive architecture search difficult and overfitting likely.Capturing long context Action recognition involves capturing spatiotemporal context across frames. Additionally, the spatial information captured has to be compensated for camera movement. Even having strong spatial object detection doesn't suffice as the motion information also carries finer details. There's a local as well as global context with respect to motion information which needs to be captured for robust predictions. Designing classification architectures Designing architectures that can capture spatiotemporal information involve multiple options which are non-trivial and expensive to evaluate. For example, some possible strategies could be

- One network for capturing spatiotemporal information vs. two separate ones for each spatial and temporal
- Fusing predictions across multiple clips
- End-to-end training vs. feature extraction and classifying separately. The most popular and benchmark datasets have been UCF101 and Sports1M for a long time. Searching for reasonable architecture on Sports1M can be extremely expensive. For UCF101, although the number of frames is comparable to ImageNet, the high spatial correlation among the videos makes the actual diversity in the training much lesser. Also, given the similar theme (sports) across both the datasets, generalization of benchmarked architectures to other tasks remained a problem. This has been solved lately with the introduction of Kinetics dataset.

V.RESULTS

The goal of ASD risk detection was modeled as a binary classification problem. The CNN was trained on 80% of the interaction data and the remaining 20% were used to validate its performance. The CNN achieved a training accuracy of 0.883 and a training loss of 0.232. Two additional machine learning classifiers (Random Forests [10] and K-Nearest Neighbor [11]) were used to situate the performance of the CNN by comparing their accuracy, precision, and recall values (Table 2).machine learning classifiers have simpler structures than a deep CNN with fewer hyper parameters that require fine-tuning. Figure 7 shows the confusion matrix resulting from the classifications generated by the CNN.

SJIF Impact Factor 2022: 8.197| ISI I.F. Value:1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 7 | Issue: 5 | May 2022 - Peer Reviewed Journal

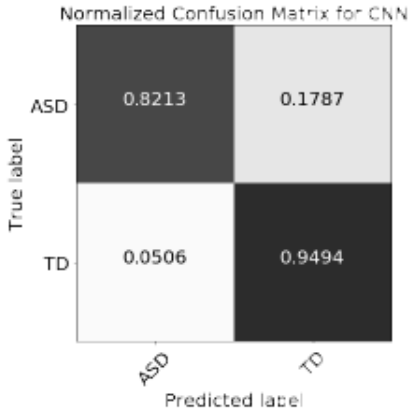

Figure 7: Confusion Matrix for ASD Action vs Ordinary Action

Figure 8: Spinning Action Identified

VI. Conclusion and Future Work

ASD is a neurological and developmental disorder that affects the interaction communication, and learning mechanism of a person's life. Numerous reports and studies have shown that treatment of early ASD individuals can cure or suppress its effects at a certain level, which might allow a person to lead a healthy life altogether. Our work analyzed the scopes and potentialities of data-driven activity analysis in the automated detection of autism. To this end, we reviewed state-of-the-art data-driven approaches for ASD detection through activity analysis, such as repetitive behavior, atypical gait patterns, and unusual visual saliency. Besides, this paper provided an analysis of different machine learning and deep learning algorithms with the results obtained in the ASD/TD detection. Moreover, possible challenges with probable solutions, available resources, and ideal experimental setups have been described briefly in this work. According to our findings, this technology has already proved its capacity to be an alternative to the traditional clinical analysis of ASD detection methods, usually taking a prolonged period with minimal certainty of the service's feasibility to the mass population. Nevertheless, some constraints may limit the accuracy and flexibility of automated detection. However, advancements in learning algorithms and computational devices will soon pave the way for more improvement
and adaptation for such data-driven approaches.

References

1. *Aly, S., Trubanova, A., Abbott, L., White, S., and Youssef, A. (2015).Vt-kfer: A kinectbased RGBD Time Dataset for Spontaneous and Non-spontaneous Facial Expression Recognition. In 2015 International Conference on Biometrics (ICB),(pp. 90–97).*
2. *American Psychiatric Association (2013). Diagnostic and Statistical Manual of Mental Disorders: DSM-5 (5th ed). American Psychiatric Association.*
3. *Arnott, B., McConachie, H., Meins, E., Fernyhough, C., Le Couteur, A., Turner, M., &Leekam, S. (2010). The Frequency of Restricted and Repetitive Behaviors in a Community Sample of 15-month-old infants. Journal of Developmental & Behavioral Pediatrics, 31(3):223-229.*
4. *Arru, G., Mazumdar, P., & Battisti, F. (2019, July). Exploiting Visual Behaviour for Autism Spectrum Disorder Identification. In 2019 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), (pp. 637-640).*
5. *Asgari, M., Bayestehtashk, A., and Shafran, I. (2013). Robust and Accurate Features for Detecting and Diagnosing Autism Spectrum Disorders. In Inter Speech, (pp. 191–194).*

SJIF Impact Factor 2022: 8.197| ISI I.F. Value:1.241| Journal DOI: 10.36713/epra2016 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)
Volume: 7 | Issue: 5 | May 2022 - Peer Reviewed Journal

6. *Avci, A., Bosch, S., Marin-Perianu, M., Marin-Perianu, R., &Havinga, P. (2010, February). Activity recognition using inertial sensing for healthcare, well-being and sports applications: A survey. In 23th International conference on Architecture of Computing Systems, (pp. 1-10).*
7. *Bai, Z., Blackwell, A. F., and Coulouris, G. (2014). Using augmented reality to elicit pretend play for children with autism. IEEE Transactions on Visualization and Computer Graphics, 21(5):598–610.*
8. *Bernardes, M., Barros, F., Simoes, M., and Castelo-Branco, M.(2015). A Serious Game with Virtual Reality for Travel Training with Autism Spectrum Disorder. In 2015 International Conference on Virtual Rehabilitation (ICVR), (pp. 127–128).*
9. *Bodfish, J. W., Symons, F. J., Parker, D. E., and Lewis, M. H.(2000b). Varieties of Repetitive Behavior in Autism: Comparisons to Mental Retardation.Journal of Autism and Developmental Disorders, 30(3):237–243.*
10. *Bodfish, J. W. (2007). Stereotypy, self-injury, and related abnormal repetitive behaviors. In Handbook of Intellectual and Developmental Disabilities, (pp. 481-505). Springer, Boston, MA.*
11. *Boucenna, S., Narzisi, A., Tilmont, E., Muratori, F., Pioggia, G., Cohen, D., &Chetouani, M. (2014). Interactive Technologies for Autistic Children: A Review. Cognitive Computation, 6(4):722-740.*
12. *Bryson, S. E., Rogers, S. J., and Fombonne, E. (2003). Autism spectrum disorders: early detection, intervention, education, and psycho-pharmacological management. The Canadian Journal of Psychiatry, 48(8):506–516.*
13. *Calhoun, M., Longworth, M., & Chester, V. L. (2011). Gait patterns in children with autism. Clinical Biomechanics, 26(2):200-206.*
14. *Capriola-Hall, N. N., Wieckowski, A. T., Swain, D., Tech, V., Aly, S.,Youssef, A., Abbott, A. L., and White, S. W. (2019). Group differences in facial emotion expression in autism: Evidence for the utility of machine classification. Behavior Therapy, 50(4):828–838.*
15. *Carcani-Rathwell, I., Rabe-Hasketh, S., and Santosh, P. J.(2006). Repetitive and stereotyped behaviours in pervasive developmental disorders. Journal of Child Psychology and Psychiatry, 47(6):573–581.*
16. *Casas, X., Herrera, G., Coma, I., & Fernández, M. (2012). A Kinect-based Augmented Reality System for Individuals with Autism Spectrum Disorders. In Grapp/Ivapp, (pp. 440-446).*
17. *Chawarska, K. and Shic, F. (2009). Looking but not seeing: A typical visual scanning and recognition of faces in 2 and 4-year-old children with autism spectrum disorder. Journal of Autism and Developmental Disorders, 39(12):16-63.*
18. *Chen, S., & Zhao, Q. (2019). Attention-Based Autism Spectrum Disorder Screening With Privileged Modality. In Proceedings of the IEEE International Conference on Computer Vision (pp. 1181-1190).*