Human Gesture and Gait Analysis for Autism Detection

Sania Zahan [1], Zulqarnain Gilani [2], Ghulam Mubashar Hassan [1] and Ajmal Mian [1]
[1] The University of Western Australia [2] Edith Cowan University
sania.zahan@research.uwa.edu.au s.gilani@ecu.edu.au
{ghulam.hassan, ajmal.mian}@uwa.edu.au

Abstract

Autism diagnosis presents a major challenge due to the vast heterogeneity of the condition and the elusive nature of early detection. Atypical gait and gesture patterns are dominant behavioral characteristics of autism and can provide crucial insights for diagnosis. Furthermore, these data can be collected efficiently in a non-intrusive way, facilitating early intervention to optimize positive outcomes. Existing research mainly focuses on associating facial and eye-gaze features with autism. However, very few studies have investigated movement and gesture patterns which can reveal subtle variations and characteristics that are specific to autism. To address this gap, we present an analysis of gesture and gait activity in videos to identify children with autism and quantify the severity of their condition by regressing autism diagnostic observation schedule scores. Our proposed architecture addresses two key factors: (1) an effective feature representation to manifest irregular gesture patterns and (2) a two-stream co-learning framework to enable a comprehensive understanding of its relation to autism from diverse perspectives without explicitly using additional data modality. Experimental results demonstrate the efficacy of utilizing gesture and gait-activity videos for autism analysis.

1. Introduction

Autism Spectrum Disorder (ASD) is a neurodevelopmental condition that poses significant communication and behavioral challenges [1]. Children with autism usually interact or behave differently compared to typically developing (TD) children. Their cognitive abilities can vary greatly, ranging from exceptional talent to severe challenges in learning, thinking, and problem-solving [19]. Whereas the reported prevalence of autism worldwide is one in 100 children, this figure is just an average, and the actual numbers can be substantially higher [3].

Autism can be detected in children as early as 18 months old or even younger [26]. Early intervention and specialized support services can substantially improve a child's development [11]. However, delayed diagnosis often results in a missed opportunity for children to receive crucial early assistance [26]. Furthermore, diagnosis can be challenging due to the lack of reliable and efficient diagnostic tools. Parents are often reluctant to accept the condition or fail to detect subtle behavioral cues [16,33]. This can lead to months of wasted time before a child gets access to proper support.

Clinically, autism is diagnosed in a face-to-face interactive session with a trained health professional who analyzes certain behavioral traits using verbal and non-verbal tasks. Communication and language assessment are essential factors in the diagnosis [17]. However, approximately 40 percent of children with autism are nonverbal [4] which further complicates the diagnosis process for this specific population. An initial diagnosis in a more suitable and accessible way is needed to facilitate higher detection accuracy of ASD. This can ensure early intervention and access to customized therapies for effective management.

Throughout the years, researchers have proposed several methods for ASD detection [5,8–10,21,29,30,32,35,38,39,41]. Many of these methods primarily focus on appearance-based features [8, 9, 21, 22, 35, 38]. However, appearance does not provide detailed insights into the autistic behavioral traits such as social-emotional exchanges, communication difficulties, stereotyped activities, etc which form a crucial part of the diagnosis [2]. Recent research shows that children with autism usually exhibit distinctive gait and gesture activity patterns [4]. Leveraging these patterns can facilitate the extraction of a distinguishable feature distribution, thus improving classification accuracy.

The unique atypical gesture activities of children with autism may include:

- **Repetitive movements** such as rocking, arm flapping, or spinning, known as "stereotypy".
- **Limited range of gestures** and may also have difficulty understanding the gestures of others.
- **Atypical gait and posture** resulting in unbalanced movement and instability in joints.

- **Impaired motor coordination** which can lead to difficulties with fine motor skills, such as writing, or gross motor skills, such as running.

In this paper, we propose a video-based method that takes a holistic view of gait and gesture to detect subtle disparities in ASD children. We evaluate the proposed method on a video dataset collected from children with and without ASD. Experimental results demonstrate the effectiveness of gesture activities in accurately identifying children with autism. This has the potential to significantly impact early diagnosis and management of ASD by providing a reliable, non-intrusive, and efficient tool for autism classification. Moreover, the action perspective facilitates the assessment of children with limited verbal communication.

The severity of autism is measured using Autism Diagnostic Observation Schedule (ADOS) score. However, there is currently no research on autism severity prediction in terms of ADOS score regression. Since the severity of autism influences gait and gesture patterns, a comprehensive analysis of activity videos can help identify subtle irregularities to assess the severity. Therefore, in this paper, we represent our analysis of ADOS score regression.

In a nutshell, our contributions are as follows.

- We propose a novel angular feature matrix which is embedded into the input skeleton and encoded using a Graph Convolutional Network (GCN). Our proposed angle embeddings enable the GCN to detect the peculiar slant in the gait posture of ASD children. To the best of our knowledge, this is the first research that focuses on this aspect of gait posture in the machine learning paradigm.
- We automatically predict ADOS scores which are highly correlated with ADOS scores measured by human experts.
- We perform a detailed analysis of the gait posture of both ASD and TD children and investigate the asymmetry in their gait.

2. Related Work

Existing research has explored various approaches for the detection of autism, with a prominent emphasis on facial expression and eye-gaze pattern-based techniques.

2.1. Facial expression and eye-gaze pattern

Physical appearance is a distinguishable feature of autism. In [38], developmental delays are detected from physical appearance in home videos. Asymmetry in facial appearance is studied in [35]. The study finds that asymmetric features are more common in people with a history of ASD. Other studies also corroborate that children with ASD display higher facial asymmetry [8, 9].

The eye-gaze pattern is also a salient marker of autism as ASD children show decreased attention than TD children [31]. Their facial expressions and eye gaze lack engagement with surrounding environments. This reduced eye-gaze pattern is stable across all ages and cultures [27]. The stacked accumulative histogram proposed in [22], captures these anomalies in eye movement trajectory. The method requires manual labeling of the eye region, and a tracking algorithm [20] to obtain the trajectories. The displacement features represent higher disparity among different visual zones for children with ASD. AttentionGazeNet [21] generates a projection of screen coordinates from 3D gaze vectors. Experiments indicate that gaze vectors are more dispersed for children with ASD. Similarly, [18] also finds a substantial difference in eye movement patterns between ASD and TD children.

2.2. Gesture pattern

In [6], a wide disparity is found in hand gesture patterns between ASD and TD children. During gameplay on a smart tablet, children with ASD used more force and gesture pressure within a greater mean area. Another gesture-based research in [41] hypothesizes that the disparity in gesture patterns in performing actions also extends to the onset, embedding information about the intention. Thus, the intended gestures can be used to diagnose children with ASD. These studies indicate the usability of motor functions in ASD analysis.

Another research direction focuses on classifying atypical actions from videos. The Bag-of-visual-words approach [30] treats image grids as visual words to recognize relevant feature descriptors. In [39], a temporal pyramid network is used to create layers of feature maps from videos with extended duration. A separate repetitive behavior discriminator is used to boost the training process by distinguishing samples with atypical actions. The atypical action classification approach in [32] uses an anchor action instance to facilitate pair-wise similarity with the target embedding. [29] analyzes action and emotion recognition from ASD therapy videos.

In [5], skeleton-based handcrafted features are used to classify ASD children. The attention-based ASD screening method in [10] exploits multiple modalities to embed complementary multimodal knowledge in a shared space.

Though extensive research has been done on appearance-related abnormalities in autism, this approach offers a limited perspective. Gesture pattern, on the other hand, provides a comprehensive perspective on physical behavior. However, existing gesture-based autism detection methods mainly use end-to-end deep learning and do not incorporate attention to the underlying mechanisms of atypical behav-

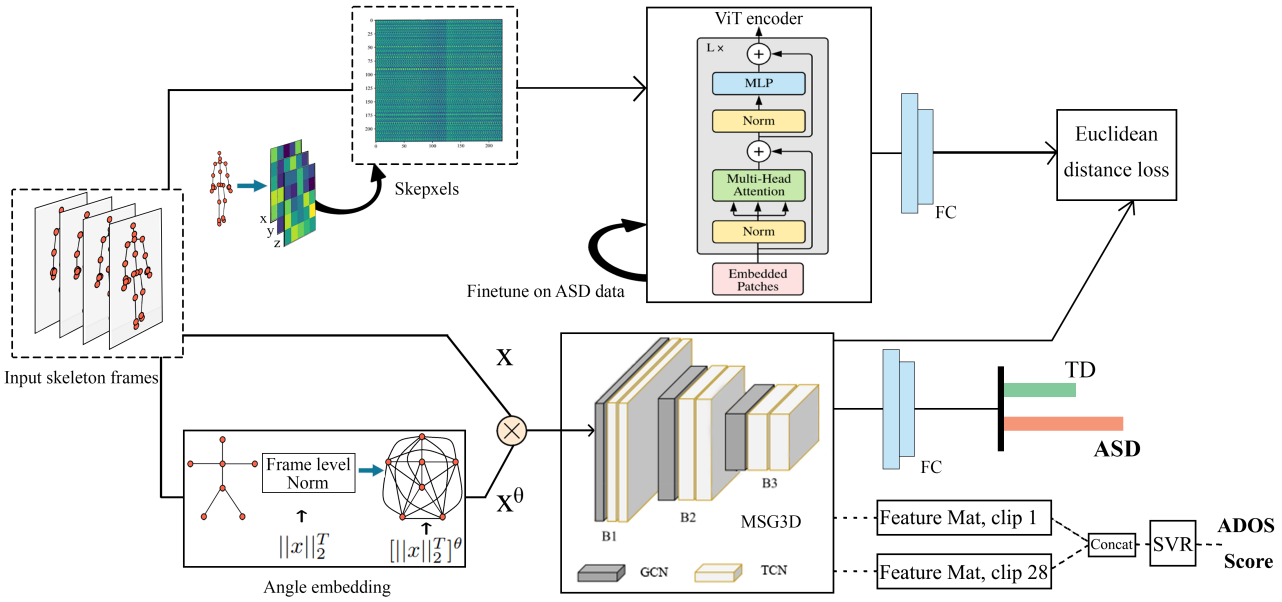


Figure 1. Proposed method: Angular feature matrix $[||x||_2^T]^\theta$ is embedded into the input skeleton ($||x||_2^T$ indicates normalization), which is then encoded by MSG3D using GCN followed by TCN. Vision Transformer (ViT) computes an aggregated embedding of the Skepxels and is used during training for pair-wise distance loss. For classification, an FC layer is used after the GCN layer. For score regression, clip-based features are extracted from the GCN and concatenated. SVR is used to predict the final ADOS scores at the video level.

iors in ASD children. Therefore, in this study, we investigate atypical gesture patterns and integrate them into the learning process to enhance the representation of anomalies.

3. Method

Our proposed method is built on the hypothesis that ASD is distinguishable solely from gesture patterns [14,28]. Skeleton-based representation facilitates visualizing gesture patterns effectively, which can be encoded into spatio-temporal embedding using graph convolution. By incorporating the joint angles, our method enhances feature representation and enables the extraction of comprehensive structural aspects and aberrations in human body movement. Overall, the skeleton frames are embedded with angle information, then encoded using MSG3D [25] and a fully connected layer (FC) generates the final mapping to the classes. For score regression, the output from MSG3D is concatenated and then used in SVR. Vision transformer is only used during training to expand the learning capacity by processing the data from a different perspective. Figure 1 illustrates our proposed method.

3.1. Graph Convolutional Network: GCN

Graph convolutional networks (GCN) [25, 40] can capture the underlying structures and encode how nodes are connected. GCN applies a localized convolutional operation to each node and its neighbors. A GCN block incorporates separate spatial (GCN) and temporal convolution (TCN) to leverage frame-wise and global attention. GCN applies filtering over the spatial dimension to encode spatial features and TCN encodes temporal features by applying filters over the temporal dimension. Multiple stacked GCN blocks generate increasingly abstract representations of the input graph.

Human skeletons can be represented by graphs as $G = (V, E)$ where $V = \{v_1, ...v_N\}$ is the set of nodes (joints) and $E = \{e_1, ..e_N\}$ is the set of edges (bones). The adjacency matrix $A \in R^{N \times N}$ represents local joint adjacency. We used MSG3D as our GCN encoder [25] to aggregate important spatio-temporal features. MSG3D for skeleton graph convolution can be represented below, as mentioned in Eq 1 in [25]

$$X_t^{(l+1)} = \sigma(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}X_t^{(l)}\theta^{(l)}), \quad (1)$$

where $\tilde{A}$ is the modified adjacency matrix, $\tilde{D}$ is the diagonal degree matrix of $\tilde{A}$, θ is the learnable weight matrix, and σ is the ReLU activation function. $\tilde{A}$ is modified A with added self-loops I and is computed by identifying the shortest k distance joints and subtracting graph powers as $\tilde{A}_{(k)} = I + \mathbb{1}(\tilde{A}^k \geq 1) - \mathbb{1}(\tilde{A}^{k-1} \geq 1)$. MSG3D uses multiple graph convolutions at different scales to extract different levels of details or resolution. Thus, it is able to efficiently capture both local and global patterns in the input graph. The encoded features are then used for ASD classification and ADOS score prediction.

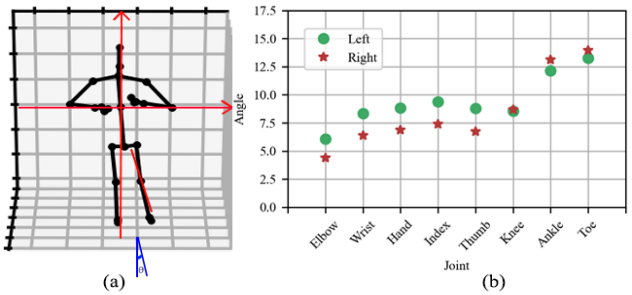


Figure 2. (a) Initial gait pose of an ASD skeleton: the red lines indicate the coordinates through the spine joint. (b) Calculated angles of the left (green) and right (red) side joints.

3.2. Angle embedding

Our statistical analysis of variations and joint distributions in gait between ASD and TD children (detail in Section 5.1) revealed that ASD children have a tendency to walk with a slanted gait posture compared to TD children. Additionally, we also found that atypicalities and asymmetry on the left and right sides of their body create irregular joint positions and movements. Due to slanted and asymmetric gait, the joints in the ASD skeleton samples form a much higher angle with the spine line. We find this to be a distinctive feature of ASD children. Figure 2 illustrates the slanted posture and calculated joint angles.

We embed the angle features into the input skeleton stream to enhance the inherent gait disorder prevalent in children with ASD. First, the input skeletons are normalized over the frame dimension. Then angle between each joint is calculated which creates a 25×25 feature matrix. Finally, this angle matrix is multiplied with the input skeleton stream over the joint dimension to generate the embedded features. Eq 2 illustrates the computation process.

$$X_{norm} = \sqrt{\sum_{t=1}^{T} |X|^2}; \quad \bar{X} = \frac{X}{X_{norm}}; \quad X^{\theta}_{i,j} = \bar{X}_i \cdot \bar{X}_j \tag{2}$$

where X is the input skeleton, $\bar{X}$ indicates L2 normalization over frame dimension T and $X^{\theta}_{i,j}$ represents the cosine angle of each joint i with all joints j where $j = 25$, calculated using dot product of each joint with all 25 joints as $\bar{X}_i \cdot \bar{X}_j = |\bar{X}_i| \cdot |\bar{X}_j| cos\theta$. This embedding process increases the distinction between ASD and TD skeleton feature space.

3.3. Skeleton Picture Elements: Skepxel

Mainstream vision models such as vision transformers (ViT) have exhibited exceptional performance, yielding remarkable results in various tasks. However, our experimental results indicate that direct use of skeleton frames in ViT produces poor performance since the joint locations are not

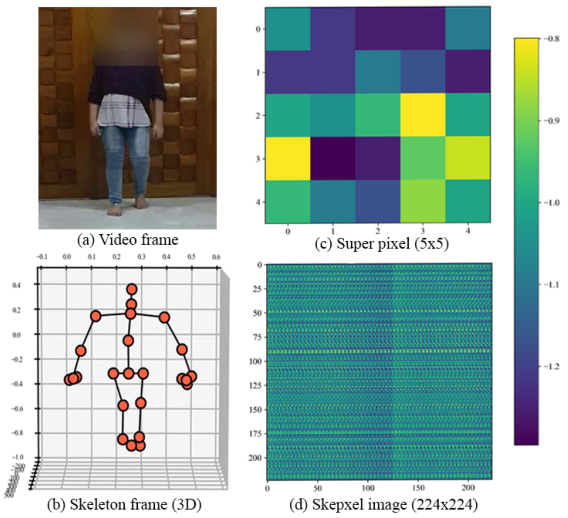


Figure 3. Illustration of (a) a video frame, (b) skeleton frame, (c) the corresponding 5x5 super pixel, and the (d) Skepxel image [23].

images. To address this limitation, we consider other representations of the skeleton frames. However, due to the limited number of joints available, the potential for generating diverse representations is constrained. The image representation of skeleton frames proposed in [13, 23] offers a comprehensive view and is suitable for processing with ViT. However, [13] focuses solely on global spatio-temporal information resulting in smaller resolution images. On the other hand, [23] takes into account both local and global spatio-temporal correlations of joints simultaneously, providing a more holistic representation of the data with higher resolution, termed as Skepxels. We leverage Skepxels to extract additional meaning from skeleton joints during training. The use of Skepxels allows us to effectively guide GCN during training to encode the complex motions and structures of different gesture patterns, leading to improved accuracy in ASD classification. Figure 3 represents a sample skeleton frame and corresponding Skepxel image.

Skepxels are constructed [23] by organizing the skeleton joints in different orders. Three channels of the joints act as the three RGB color channels. Super pixels in a single column represent different ordered joints of the same frame and each row contains superpixels from different temporal frames. This systematic construction results in an aggregated image representation that encodes heterogeneous semantic perceptions. For further details read the paper [23].

3.4. Vision transformer encoder: ViT

As outlined in Figure 1, we use ViT [12] to encode aggregated feature representation of Skepxels. ViT takes non-overlapping patches of the image and projects them onto a feature vector through a learnable linear projection. It uses positional embedding to retain the order of the patches. The self-attention mechanism allows it to selectively attend to

relevant parts of the image, enabling it to capture global context more effectively than conventional CNNs. We use an MLP layer after ViT to map the Skepxel embedding to the same feature space as the GCN stream.

During the training phase, the Skepxel embedding is incorporated into the model training to facilitate co-learning of the spatial and temporal features. Euclidean distance loss between the Skepxel embedding and joint embedding from GCN is added with the classification loss. However, it is not used during the testing phase, where the model relies solely on the learned representations from the GCN stream.

4. Datasets

We evaluate our proposed method on two datasets. We use the Gait and Full Body Movement dataset for ASD classification [5] and the DREAM dataset for ADOS score regression [7].

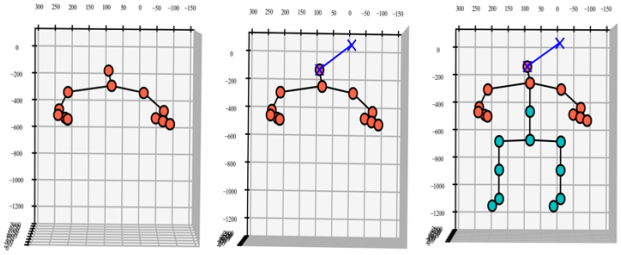
4.1. Gait and Full Body Movement dataset

This dataset used Kinect v2 and Samsung note 9 to collect 3D joint coordinates (skeleton videos) and RGB videos [5]. The dataset was collected in a controlled environment where children walked 2.5m (approximately two gait cycles) in front of the camera ten times. Then one gait cycle was extracted from an eligible candidate of the ten trials. Finally, faces were detected using Haar Cascade or MTCNN and blurred with a Gaussian filter to obscure the identities for anonymity in the RGB videos. The dataset contains fifty-nine children with ASD (9 of these children have severe ASD) and fifty TD children. There are a total of 109 samples. The dataset also contains seven augmented versions of each sample using jittering, scaling, left and right translation, horizontal and vertical flipping, and slicing, increasing the augmented dataset size to 700 samples.

We follow two approaches for train and test sample selection: random shuffling of subjects, mentioned as random data and sliding window to select blocks of subject range, mentioned as block data in the results section. With block data sample selection, we ensure that each subject is evaluated at least once as either a training or testing sample. Furthermore, whenever a subject is selected for either training or testing, both original and augmented versions go to the same split to ensure no leakage of the augmented samples to the testing set or vice versa.

4.2. DREAM dataset

The DREAM dataset incorporated 61 children (9 female) aged between 3 to 6 years with varied levels of autism and ADOS scores ranging from 7-20 [7]. The samples were collected in a therapy environment where the children interacted with either a human therapist (SHT) or a humanoid robot (RET). They performed three different tasks: imitation, joint attention, and turn-taking. The sessions were conducted in the following manner. The interaction partner (human or robot) provided a discriminative stimulus and waited for the response from the child. Positive feedback or indication to repeat was provided depending on the child's behavior. Each session duration varied from 3 to 87 minutes, with a median duration of 32 minutes.



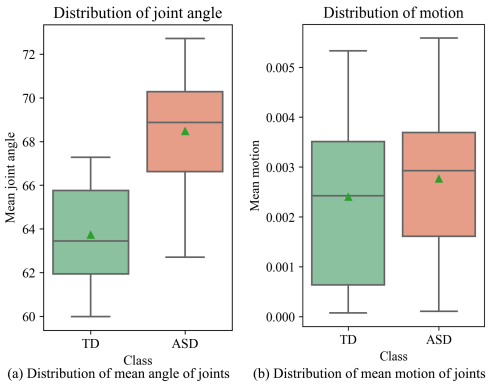
(a) Upper body (b) Upper body with Eye gaze (c) Full body

Figure 4. Sample skeleton frame from DREAM dataset (a) original upper body skeleton from the dataset (b) upper body skeleton with eye gaze as additional head joint and (c) full body skeleton with interpolated lower joints.

The dataset provides ADOS scores from two diagnosis sessions, initial and intervention. We consider only the initial diagnosis samples for this work as final ADOS scores are not published for the intervention sessions. ADOS score is a semi-structured autism diagnosis technique that is most commonly used to measure the severity of ASD. In addition to a raw total score, it includes individual scores for several skill factors such as communication, age, language level, interaction, etc., to categorize children into homogeneous groups [15]. Based on their age and the ADOS score evaluation module used, children with ASD can be divided into three classes: NonSpectrum (NS), Autism Spectrum Disorder (ASD), and Autistic (AUT, the most severe case of autism), following the metrics provided in [15]. We design our work to predict the ADOS score and classification of the samples into the above three classes based on the predicted scores. Figure 4 shows a sample skeleton.

4.3. Data preprocessing

We redesign the skeletons in the DREAM dataset as it contains only upper-body joints. The missing joints are interpolated from the existing joints in the different direction. Thus the incomplete 10 joints skeleton structure is converted to a full-body structure with 25 joints. This enables us to process these skeletons using our proposed method where the GCN module expects the input to be 25 skeleton joints. We use the eye-gaze vector as the third head joint. This approach adds an extra perspective to our modified 25-joint skeletons by integrating the subject's visual attention. We replace any missing eye-gaze values with preceding values. The dataset is preprocessed to have a view-invariant transformation with the shoulder joints aligned with the x-axis and spine joints aligned with the z-axis. The spine joint is translated to the origin (0,0,0). We repeated frames where



(a) Distribution of mean angle of joints (b) Distribution of mean motion of joints

Figure 5. (a) Distribution of joint angle, higher median indicates ASD children have a much higher median joint angle. (b) Distribution of motion of TD and ASD samples; a larger box of TD samples demonstrates that TD children express higher variations in motion while walking.

necessary to maintain a fixed-length video.

For the Gait and Full Body Movement dataset, we do not perform rotation as it will eliminate the slanted gait posture of ASD skeletons.

5. Experiments

Our statistical analysis of the gait pattern reveals discriminative atypicalities in ASD samples and our proposed method exploits this insight for autism classification and ADOS score regression.

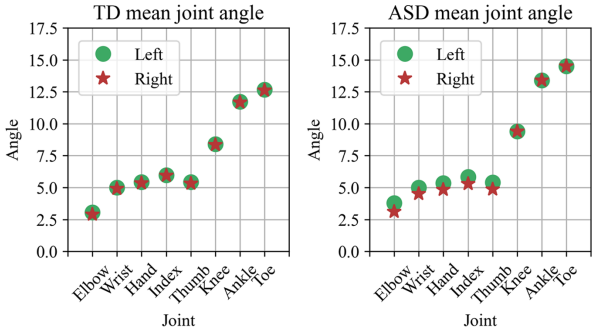
5.1. Statistical analysis

In this section, we present a comparative statistical analysis between ASD and TD samples. ASD samples exhibited distinctive variations and asymmetry over the entire population.

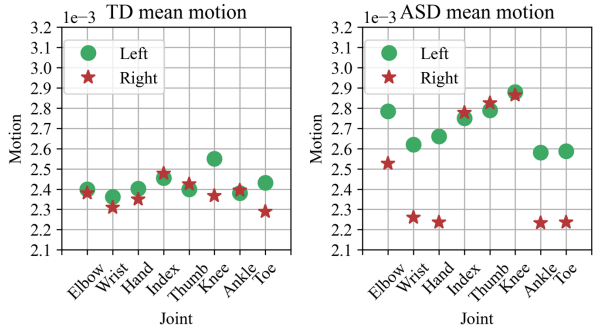
Variation in joint and motion distribution: Children with ASD represent restricted and repetitive behavioral patterns [9]. They usually tend to have a slower gait and may have visible difficulty while walking. These atypical gaits create a slanted posture. We analyzed the angle of each joint with the spine (center) joint in each frame and calculate the mean. Figure 5 (a) illustrates the distribution of these joint angles for TD and ASD children. ASD children present a much higher distribution than TD children, demonstrating a higher mean joint angle.

Figure 5 (b) shows the distribution of mean motion. The limited range of gestures and slower gait cycle of ASD children means that the distribution of their movement pattern will be less dispersed. On the other hand, TD children usually show comparatively more diverse motion, resulting in a broader motion distribution.

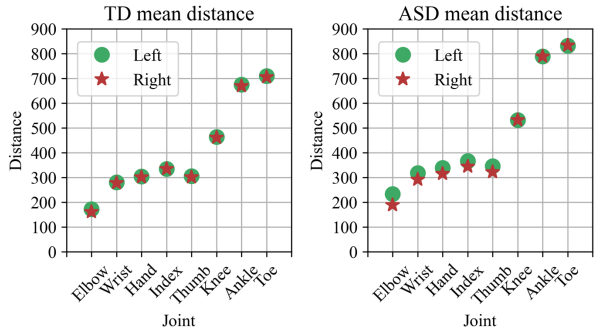
Asymmetry in gait: Recent studies have revealed that children diagnosed with ASD exhibit hypermasculine traits and asymmetry in their facial morphology [8, 9, 34–37]. Based on the atypical gesture behaviors observed in children with ASD, we hypothesize that this asymmetry may also manifest in their walking patterns or gait. In order to ascertain whether there is an asymmetry in the gait, we compared the angle, motion, and distance between joints on the left and right sides of the body. Our observations are illustrated in Figure 6.



(a) Comparison between mean angle of left and right joints with spine (center joint)



(b) Comparison between mean motion of left and right joints



(c) Comparison between mean distance of left and right joints from spine (center joint)

Figure 6. Illustration of asymmetry in the joints of the left and right sides of the body. (a) Comparison between mean joint angle (in degrees): left and right side joints are overlapped in TD samples whereas deviated in ASD samples. (b) Comparison between mean motion. ASD samples indicate much higher as well as asymmetric movement patterns in left and right side joints. (c) Mean distance (in centimeters) of the left and right side joints from the spine is depicted. Distance in ASD samples is more dispersed which represents that ASD children have asymmetric hand and leg positions during walking.

A typical person tends to have uniform hand and leg movement coordination in a complete gait cycle. A higher difference in angle, motion, and distance between the left and right side joints indicates higher asymmetry. We select 16 joints, including 10 hand joints (left and right) and 6 leg joints (left and right) and calculate the angle, motion, and distance of each joint with the corresponding spine joint in the same frame. As can be seen in Figure 6 (a), (b), and (c), ASD samples depict higher differences in the values between left and right joints than TD. Thus, we can conclude that the gait of children with ASD expresses higher asymmetry.

5.2. Quantitative results

Experimental results on the datasets for ASD detection and ADOS score regression are discussed in the following sections.

Results on Gait and Full Body Movement dataset: Table 1 presents average results from 10-fold cross-validation along with standard deviation on the Gait and Full Body Movement dataset. In the Table, **MSG3D** indicates results for the initial model proposed in [25]. We used pretrained weights of MSG3D trained on NTU RGB+D 120 action recognition dataset [24]. **Angle infusion** represents joint angle embedding into the input skeletons, and **Skepxel distance loss** indicates two-stream training where the Skepxels are used for pair-wise distance loss calculation. **NoAug** specifies evaluation on the original test samples only.

Table 2 illustrates the comparison between our proposed method and existing work on the Gait and Full Body Movement dataset. Ahmed et al. [5] did not share the train-test subject split and presented a single experimental result. We perform more rigorous experiments with 10-fold cross-validation, ensuring that every subject ends up in the test set. Our proposed method achieves 93% accuracy on average on the 10-folds, with a minimum of 86.67% and a maximum of 100% accuracy.

Results on DREAM dataset: The DREAM dataset contains only ASD samples with corresponding ADOS scores and ADOS-related information. Samples in this dataset are way longer than the Gait and Full Body Movement dataset, with a maximum number of frames of around 60K. Therefore, we finetune our model on this dataset using samples with one minute duration. Then the trained model is used to extract features from the other frames of all samples. Finally, we use Support Vector Regression (SVR) to predict the ADOS scores.

We calculate the corresponding classes using the predicted ADOS scores and the corresponding ADOS module and age following the metrics provided in [15]. NS class includes the following population: for ADOS module 1 - ages ($\geq$3 and $\leq$6 years) with scores $\leq$10 and for module 2 - ages (3 and 4) with scores (6 and 7) and ages (5 and 6) with scores $\leq$6 [15]. ASD class includes: for module 1 - ages $\geq$6 with scores ($>$10 and $\leq$15) and for module 2 - ages (3 and 4) with scores ($>$6 and $\leq$9) and ages (5 and 6) with scores =8 [15]. AUT class includes: for module 1 - ages ($\geq$3 and $\leq$6) with scores $>$15 and for module 2 - ages (3 and 4) with scores $>$9 and ages (5 and 6) with scores $>$8 [15]. Table 3 represents the error rates for score regression, spearman correlation (SP), *P* values, and classification accuracy calculated from the predicted scores. Spearman correlation and *P* values indicate that the predicted ADOS scores are highly correlated with the actual scores measured by expert professionals.

Subjects in the DREAM dataset perform three tasks: Turn taking, Imitation, and Joint attention. We analyze the overall and the per-task accuracy and achieve 78.6% average accuracy with individual tasks. Figure 7 illustrates the classification accuracy per ADOS scores for all tasks.

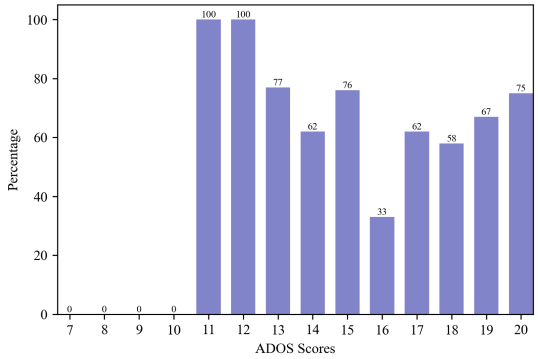


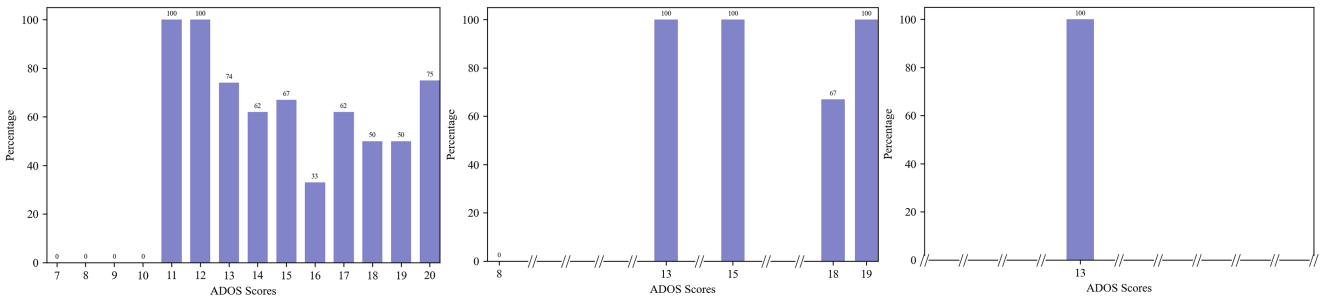
Figure 7. Classification accuracy per ADOS scores for all tasks.

Figure 8 represents classification accuracy for individual tasks. For tasks 1 and 2, the model fails to correctly classify ADOS total in the range of 7-10. This ADOS score range is the hardest as they fall within the NonSpectrum class, where children do not have autism but exhibit very subtle atypical behaviors.

6. Discussion

The gait and full body movement dataset comprises single gait cycles, with a restricted number of frames. Due to their brevity, anomalous behavioral expressions exhibited by children with ASD are scant in these sequences. However, our statistical analysis extracted valuable insights into the gait and gesture traits. The observed gait asymmetry and increased joint angle, motion, and distance in standing pose can provide a significant attribute for autism analysis. Longitudinal data with extended duration will provide a better understanding of autistic gesture expressions. It will facilitate faster diagnosis and a more personalized therapy and support system.

Experimental results on the DREAM dataset indicate that our model fails to classify samples with lower ADOS



(a) Classification accuracy per ADOS score for task 1 (b) Classification accuracy per ADOS score for task 2 (c) Classification accuracy per ADOS score for task 3

Figure 8. Classification accuracy per ADOS scores for individual tasks (a) task 1 - Turn taking (b) task 2 - Imitation and (c) task 3 - Joint attention.

	Accuracy							
	Block data				**Random data**			
	MSG3D	+Angle infusion	+Skepxel distance loss	NoAug	MSG3D	+Angle infusion	+Skepxel distance loss	NoAug
Avg	86.38	**89.10**	**90.86**	**93.00**	91.24	**92.66**	**91.48**	**92.33**
± SD	± 5.48	**± 6.67**	**± 4.84**	**± 4.82**	± 3.19	**± 3.01**	**± 3.90**	**± 3.67**

Table 1. Average results with standard deviation from ten fold cross-validation on the Gait dataset. The results clearly demonstrate that the inclusion of additional modules in our proposed architecture yields notable improvements in comparison to the baseline model: MSG3D.

Method	**Data**	**Subject selection**	**Accuracy**
Ahmed et al. [5]	skeleton	Random	92.00
Ours	skeleton	Random (Max)	**96.67**
		Random (Avg)	**92.33**
		Block (Max)	**100.00**
		Block (Avg)	**93.00**

Table 2. Comparison between our results and existing work on the Gait and Full Body Movement dataset.

	Error rate	**SP**	***P***	**Accuracy**
Avg	2.91	0.34	.002	51.56
± SD	± 0.27	± 0.07	± .003	± 5.10

Table 3. SVR Regression for 10-fold cross-validation. Accuracy represents % of correct classification calculated using predicted ADOS Scores.

scores ranging from 7 to 10. This range constitutes the non-Spectrum class, where subjects have mild ASD-like behavioral patterns with very few atypical traits. Additionally, the dataset incorporates a limited number of samples with no visible distinction among samples from different classes. On the other hand, the methodologies for measuring ADOS scores are sub-standardized, adding further complexity to the prediction process. We use a range of tolerance values to mitigate this issue while associating predicted ADOS scores with classes. Despite these challenges, our proposed methodology demonstrates superior performance, particularly for higher ADOS scores within the ASD and AUT classes. These two classes show more physical and behavioral anomalies and require substantial assistance and support systems. Our proposed method offers a feasible assessment solution for such cases.

Moreover, privacy concerns have significantly impeded progress in autism research. Employing a skeleton video-based assessment system can alleviate these concerns and enable a more comprehensive approach to autism analysis. Our present study opens a new research area to anonymize and automate the tedious process of autism diagnosis and ADOS score prediction.

7. Conclusion

This paper presents a comprehensive analysis of autism spectrum disorder using gait and gesture-based approaches from skeleton videos. Our statistical analysis suggests that children with ASD display asymmetrical gait patterns and higher mean joint angle distributions. These findings align with the overall trend in atypical behavioral patterns observed in individuals with ASD. Our proposed early angle embedding technique enhances GCN performance by emphasizing atypical gesture patterns with greater precision to extract robust spatio-temporal features. Moreover, we leverage the concept of "Skepxels" to enable multi-modal training without the need for supplementary input modalities. Our experimental results indicate that utilizing skeleton data holds significant potential for advancing our understanding of autism related behaviours and could serve as a promising avenue for future research.

8. Acknowledgement

Professor Ajmal Mian is the recipient of an Australian Research Council Future Fellowship Award (project number FT210100268) funded by the Australian Government. Sania Zahan is the recipient of a University Postgraduate Award and University of Western Australia International Fee Scholarship.

References

[1] What is autism spectrum disorder? https://www.cdc.gov/ncbddd/autism/facts.html, 2020. [Online; accessed 26-January-2023]. 1

[2] Autism spectrum disorder (ASD). https://www.healthdirect.gov.au/autism, 2022. [Online; accessed 05-March-2023]. 1

[3] Autism spectrum disorders. https://www.who.int/news-room/fact-sheets/detail/autism-spectrum-disorders, 2022. [Online; accessed 25-Fabruary-2023]. 1

[4] Autism statistics and facts — autism speaks. https://www.autismspeaks.org/autism-statistics-asd, 2022. [Online; accessed 26-January-2023]. 1

[5] Ahmed A. Al-Jubouri, Israa Hadi Ali, and Yasen Rajihy. Gait and full body movement dataset of autistic children classified by rough set classifier. *Journal of Physics: Conference Series*, 2021. 1, 2, 5, 7, 8

[6] Anna Anzulewicz, Krzysztof Sobota, and Jonathan T. Delafield-Butt. Toward the autism motor signature: Gesture patterns during smart tablet gameplay identify children with autism. In *Scientific Reports*, 2016. 2

[7] Erik Billing, Tony Belpaeme, Haibin Cai, Hoang-Long Cao, Anamaria Ciocan, Cristina Costescu, Daniel David, Robert Homewood, Daniel Hernandez Garcia, and Pablo et al. Gómez Esteban. The DREAM Dataset: Supporting a data-driven study of autism spectrum disorder and robot enhanced therapy. *PLOS ONE*, 2020. 5

[8] Maryam Boutrus, Syed Zulqarnain Gilani, Gail A. Alvares, Murray T. Maybery, Diana Weiting Tan, Ajmal Mian, and Andrew J. O. Whitehouse. Increased facial asymmetry in autism spectrum conditions is associated with symptom presentation. *Autism Research*, 2019. 1, 2, 6

[9] Maryam Boutrus, Zulqarnain Gilani, Murray T. Maybery, Gail A. Alvares, Diana W. Tan, Peter R. Eastwood, Ajmal Mian, and Andrew J. O. Whitehouse. Brief report: Facial asymmetry and autistic-like traits in the general population. In *Journal of Autism and Developmental Disorders*, 2021. 1, 2, 6

[10] Shi Chen and Qi Zhao. Attention-based autism spectrum disorder screening with privileged modality. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019. 1, 2

[11] National Research Council. *Educating Children with Autism*. The National Academies Press, Washington, DC, 2001. 1

[12] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, 2021. 4

[13] Yong Du, Yun Fu, and Liang Wang. Skeleton based action recognition with convolutional neural network. In *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*, 2015. 4

[14] E. Fourie, E. R. Palser, J. J. Pokorny, M. Neff, and S. M. Rivera. Neural processing and production of gesture in children and adolescents with autism spectrum disorder. *Frontiers in Psychology*, 2020. 3

[15] Katherine Gotham, Andrew Pickles, and Catherine Lord. Standardizing ADOS scores for a measure of severity in autism spectrum disorders. *Journal of autism and developmental disorders*, 2009. 5, 7

[16] Paula Grogan, Maya Yaari, Rachel Jellett, Katy Unwin, and Cheryl Dissanayake. Parent resolution of diagnosis and intervention fidelity in a parent-delivered intervention for preschool children with autism: A mixed methods study. *Research in Autism Spectrum Disorders*, 2023. 1

[17] Kristelle Hudry, Jodie Smith, Sarah Pillar, Kandice J. Varcin, Catherine A. Bent, Maryam Boutrus, Lacey Chetcuti, Alena Clark, Cheryl Dissanayake, Teresa Iacono, Lyndel Kennedy, Alicia Lant, Jemima Robinson Lake, Leonie Segal, Vicky Slonims, Carol Taylor, Ming Wai Wan, Jonathan Green, and Andrew J. O. Whitehouse. The utility of natural language samples for assessing communication and language in infants referred with early signs of autism. *Research on Child and Adolescent Psychopathology*, 2023. 1

[18] Ming Jiang and Qi Zhao. Learning visual attention to identify people with autism spectrum disorder. In *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017. 2

[19] Karen R. Johnson. Using a strengths-based approach to improve employment opportunities for individuals with autism spectrum disorder. In *New Horizons in Adult Education & Human Resource Development*, 2022. 1

[20] Zdenek Kalal, Krystian Mikolajczyk, and Jiri Matas. Tracking-learning-detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012. 2

[21] Jing Li, Zejin Chen, Yihao Zhong, Hak-Keung Lam, Junxia Han, Gaoxiang Ouyang, Xiaoli Li, and Honghai Liu. Appearance-based gaze estimation for ASD diagnosis. *IEEE Transactions on Cybernetics*, 2022. 1, 2

[22] Jing Li, Yihao Zhong, Junxia Han, Gaoxiang Ouyang, Xiaoli Li, and Honghai Liu. Classifying asd children with LSTM based on raw videos. *Neurocomputing*, 2020. 1, 2

[23] Jian Liu, Naveed Akhtar, and Ajmal Mian. Skepxels: Spatio-temporal image representation of human skeleton joints for action recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2019. 4

[24] Jun Liu, Amir Shahroudy, Mauricio Perez, Gang Wang, Ling-Yu Duan, and Alex C. Kot. Ntu rgb+d 120: A large-scale benchmark for 3D human activity understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020. 7

[25] Ziyu Liu, Hongwen Zhang, Zhenghao Chen, Zhiyong Wang, and Wanli Ouyang. Disentangling and unifying graph convolutions for skeleton-based action recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020. 3, 7

[26] Catherine Lord, Susan Risi, Pamela S DiLavore, Cory Shulman, Audrey Thurm, and Andrew Pickles. Autism from 2 to 9 years of age. *Archives of general psychiatry*, 2006. 1

[27] Xue'er Ma, Haixia Gu, and Jingjing Zhao. Atypical gaze patterns to facial feature areas in autism spectrum disorders reveal age and culture effects: A meta-analysis of eye-tracking studies. In *Autism Research*, 2021. 2

[28] A. de Marchena, E.S. Kim, A. Bagdasarov, J. Parish-Morris, B.B. Maddox, E.S. Brodkin, and R.T. Schultz. Atypicalities of gesture form and function in autistic adults. *Journal of Autism and Developmental Disorders*, 2019. 3

[29] Elisabeta Marinoiu, Mihai Zanfir, Vlad Olaru, and Cristian Sminchisescu. 3D human sensing, action and emotion recognition in robot assisted therapy of children with autism. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018. 1, 2

[30] Farhood Negin, Baris Ozyer, Saeid Agahian, Sibel Kacdioglu, and Gulsah Tumuklu Ozyer. Vision-assisted recognition of stereotype behaviors for early diagnosis of autism spectrum disorders. *Neurocomputing*, 2021. 1, 2

[31] D Riby and P J B Hancock. Looking at movies and cartoons: eye-tracking evidence from williams syndrome and autism. In *Journal of Intellectual Disability Research : JIDR*, 2009. 2

[32] Alberto Sabater, Laura Santos, Jose Santos-Victor, Alexandre Bernardino, Luis Montesano, and Ana C. Murillo. One-shot action recognition in challenging therapy scenarios. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2021. 1, 2

[33] Amy Sudhinaraset and Alice Kuo. Parents' perspectives on the role of pediatricians in autism diagnosis. *Journal of Autism and Developmental Disorders*, 2013. 1

[34] Diana Weiting Tan, Syed Zulqarnain Gilani, Gail A. Alvares, Ajmal Mian, Andrew J. O. Whitehouse, and Murray T. Maybery. An investigation of a novel broad autism phenotype: increased facial masculinity among parents of children on the autism spectrum. In *Proceedings of the Royal Society B.*, 2022. 6

[35] Diana Weiting Tan, Syed Zulqarnain Gilani, Maryam Boutrus, Gail A Alvares, Andrew J O Whitehouse, Ajmal Mian, David Suter, and Murray T Maybery. Facial asymmetry in parents of children on the autism spectrum. *Autism Res*, 2021. 1, 2, 6

[36] Diana Weiting Tan, Syed Zulqarnain Gilani, Murray T. Maybery, Ajmal Mian, Anna Hunt, Mark Walters, and Andrew J. O. Whitehouse. Hypermasculinised facial morphology in boys and girls with autism spectrum disorder and its association with symptomatology. In *Scientific Reports*, 2017. 6

[37] Diana Weiting Tan, Murray T. Maybery, Syed Zulqarnain Gilani, Gail A. Alvares, Ajmal Mian, David Suter, and Andrew J. O. Whitehouse. A broad autism phenotype expressed in facial morphology. In *Translational Psychiatry*, 2020. 6

[38] Qandeel Tariq, Scott Lanyon Fleming, Jessey Nicole Schwartz, Kaitlyn Dunlap, Conor Corbin, Peter Washington, Haik Kalantarian, Naila Z Khan, Gary L Darmstadt, and Dennis Paul Wall. Detecting developmental delay and autism through machine learning models using home videos of Bangladeshi children: Development and validation study. *Journal of medical Internet research*, 2019. 1, 2

[39] Yuan Tian, Xiongkuo Min, Guangtao Zhai, and Zhiyong Gao. Video-based early ASD detection via temporal pyramid networks. In *2019 IEEE International Conference on Multimedia and Expo (ICME)*, 2019. 1, 2

[40] Sijie Yan, Yuanjun Xiong, and Dahua Lin. Spatial temporal graph convolutional networks for skeleton-based action recognition. In *AAAI*, 2018. 3

[41] Andrea Zunino, Pietro Morerio, Andrea Cavallo, Caterina Ansuini, Jessica Podda, Francesca Battaglia, Edvige Veneselli, Cristina Becchio, and Vittorio Murino. Video gesture analysis for autism spectrum disorder detection. In *2018 24th International Conference on Pattern Recognition (ICPR)*, 2018. 1, 2