







## Towards Automated Motor Impulsivity Monitoring in Real-world Scenarios: A Multiple Object Tracking Approach

Fahmy F. Dalimarta<sup>\*1,2</sup> , Pulung N. Andono<sup>1</sup> , Moch. A. Soeleman<sup>1</sup> , Zainal. A. Hasibuan<sup>1</sup> 

<sup>1</sup> Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang 50131, Indonesia

<sup>2</sup> Faculty of Technic, Universitas Muhammadiyah Tegal, Tegal 52124, Indonesia

\*Corresponding Author: [fahmy@mhs.dinus.ac.id](mailto:fahmy@mhs.dinus.ac.id)

### ARTICLE INFO

#### Article history:

Received 12 June 2024

Revised 24 June 2024

Accepted 27 June 2024

Available online 30 January 2025

E-ISSN: 2580-829X

P-ISSN: 2580-6769

#### How to cite:

Fahmy F. Dalimarta and Prof. D. M. Sable, "Towards Automated Motor Impulsivity Monitoring in Real-world Scenarios: A Multiple Object Tracking Approach" Data Science: Journal of Computing and Applied Informatics, vol. V9, no. 1, Jan. 2025, doi: 10.32734/jocai.v9.i1-16686

### ABSTRACT

Assessment of motor impulsivity often faces several challenges. Conventional assessments that rely on controlled settings often fail to capture impulsive behaviors in real-world contexts. This study proposes an automated approach using Multiple Object Tracking (MOT) technology to assess motor impulsivity. The aim was to develop a system for detecting and quantifying motor impulsivity in naturalistic, multi-person environments. By employing cutting-edge MOT algorithms, the solution tracks multiple individuals concurrently, enabling movement and interaction analyses. This methodology integrates MOT with behavioral models to identify motor impulsivity patterns such as abrupt trajectory changes or impulsive gesturing. Trained on real-world annotated datasets, the system ensures adaptability across settings. Our approach successfully distinguished impulsive movements from typical behavioral patterns, with an accuracy of 95.43%. This approach could revolutionize assessments by providing objective and quantitative measurements and facilitating enhanced diagnostics and personalized interventions. Extensive evaluations are required to assess real-time capabilities, robustness in occluded environments, and accurate impulsive pattern identification. These findings could enable broader clinical, research, and behavioral monitoring applications, advancing our understanding of the implications of motor impulsivity.

**Keyword:** Multiple Object Tracking, Motor Impulsivity, Quantification.

### ABSTRAK

Penilaian impulsivitas motorik seringkali menghadapi beberapa tantangan. Penilaian konvensional yang mengandalkan pengaturan terkendali sering gagal menangkap perilaku impulsif dalam konteks dunia nyata. Studi ini mengusulkan pendekatan otomatis menggunakan teknologi Multiple Object Tracking (MOT) untuk menilai impulsivitas motorik. Tujuannya adalah mengembangkan sistem untuk mendeteksi dan mengukur impulsivitas motorik dalam lingkungan naturalistik dengan banyak subjek. Dengan menggunakan algoritma MOT mutakhir, solusi ini melacak beberapa individu secara bersamaan, memungkinkan analisis gerakan dan interaksi. Metodologi ini mengintegrasikan MOT dengan model perilaku untuk mengidentifikasi pola impulsivitas motorik seperti perubahan lintasan mendadak atau gerakan impulsif. Dilatih pada dataset teranotasi dunia nyata, sistem ini memastikan adaptabilitas di berbagai pengaturan. Pendekatan kami berhasil membedakan gerakan impulsif dari pola perilaku tipikal, dengan akurasi 95.43%. Pendekatan ini dapat merevolusi penilaian dengan menyediakan pengukuran objektif dan kuantitatif serta memfasilitasi diagnostik yang lebih baik dan intervensi personal. Evaluasi ekstensif diperlukan untuk menilai kemampuan real-time, ketangguhan dalam lingkungan terhalang, dan identifikasi pola impulsif yang akurat. Temuan ini dapat memungkinkan aplikasi klinis, penelitian, dan pemantauan perilaku yang lebih luas, meningkatkan pemahaman kita tentang implikasi impulsivitas motorik.

**Keyword:** Multiple Object Tracking, Impulsivitas Motorik, Kuantifikasi.



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International.

<http://doi.org/10.32734/jocai.v9i1.16686>

## 1. Introduction

The ability to regulate bodily actions and impulses is a critical aspect of cognitive and behavioral development. However, this self-regulatory capacity is compromised in some individuals, leading to a condition known as motor impulsivity. It is characterized by a tendency to act without forethought or consideration, particularly in terms of physical actions. It involves making quick, spontaneous movements without fully assessing the potential consequences. Motor impulsivity is often observed in conjunction with several neuropsychiatric and neurodevelopmental disorders, including Attention-Deficit/Hyperactivity Disorder (ADHD) [1], [2], bipolar disorder [3], and certain depressive disorders [4], [5]. Children with ADHD are characterized by prominent symptoms of motor impulsivity, where affected individuals frequently struggle to remain still and engage in disruptive behaviors that can impede social interactions and academic performance [6]. The consequences of motor impulsivity extend beyond social and educational domains, as uncontrolled motor behaviors can also pose safety risks. The prevalence of injuries among school-aged children with ADHD ranges from 3% to 7% [7], and these individuals are nearly twice as likely to sustain injuries as their neurotypical peers.

Although motor impulsivity can significantly impact various aspects of life, accurate and objective assessment of this complex phenomenon remains challenging. Traditional assessment methods [8], [9], [10] that often rely on subjective observations or controlled settings may fail to capture the nuances and real-world manifestations of impulsive motor behaviors. Accurate assessment of motor impulsivity is crucial for several reasons. First, the early and accurate identification of motor impulsivity can facilitate timely intervention, potentially mitigating adverse effects on academic, social, and overall well-being. Additionally, a deeper understanding of the underlying neurobiological and environmental factors that contribute to motor impulsivity could pave the way for more targeted and effective therapeutic approaches. Such approaches could benefit not only individuals with ADHD, but also those with related conditions characterized by impulsive motor behaviors.

Quantitative techniques have shown promise in objectively assessing and characterizing motor impulsivity [11]. Among these approaches, motion tracking and analysis using specialized equipment such as cameras or inertial sensors have been explored. These methods allow for precise measurement and quantification of various movement features, including velocity, acceleration, and directional changes. Marker-based 3D motion capture systems, which rely on optical tracking of body-fixed reflective markers, are considered the clinical reference standard for human movement analysis [12]. Researchers have used this technique to analyze aspects of limb movement, such as range of motion, velocity, and coordination, in individuals with impulsivity-related disorders. Wearable inertial sensors, including accelerometers [13] and gyroscopes [14], have also been employed to track and quantify body movements directly. These sensors can be incorporated into specialized attire or attached to specific body segments to provide precise measurements of linear and angular kinematic data such as acceleration profiles, jerk (rate of change of acceleration), and movement trajectories.

Almost all of the quantitative techniques mentioned above encounter limitations when assessing motor impulsivity in multiperson settings. It is difficult, or sometimes expensive, for optical motion capture and wearable sensor systems to accurately capture and differentiate the movements of multiple people at the same stage. This limitation is significant because many real-world scenarios involve interactions and dynamic environments involving multiple individuals. Addressing this gap is crucial as it could lead to more accurate and comprehensive assessments, enabling better understanding and interventions for impulsivity-related disorders. Overcoming this limitation requires the development of advanced motion tracking systems capable of reliably tracking and differentiating the movements of multiple subjects simultaneously, even in the presence of occlusions and overlapping movements.

The proposed approach combines the strengths of two state-of-the-art computer vision models: You Only Look Once (YOLO) and ByteTrack [15]. YOLO is a renowned object-detection algorithm that rapidly and accurately identifies multiple objects, including individuals, within a given frame. On the other hand, ByteTrack is a highly efficient and reliable multi-object tracking algorithm capable of tracking multiple individuals across consecutive video frames, even in the presence of occlusions and complex movements [16]. By integrating these two powerful models, our system can reliably detect and track the movements of multiple individuals simultaneously, thereby enabling the extraction of precise kinematic data. These comprehensive motion data served as the foundation for the subsequent analysis and classification stages.

To effectively model and classify the intricate patterns of motor impulsivity, we employed Bi-Directional Long Short-Term Memory (Bi-LSTM) [17] networks known for their ability to capture temporal dependencies and relationships in sequential data. Bi-LSTM has demonstrated remarkable success in various applications,

including gesture recognition [18], [19], [20], [21], video analysis [22], [23], and behavior detection [24], [25], [26], [27], making it well suited for the precise prediction and classification of motor impulsivity.

Our proposed multi-person detection and tracking approach, combined with the powerful sequence-modeling capabilities of Bi-LSTM networks, addresses a crucial limitation of existing methods by enabling accurate and objective assessments of motor impulsivity in complex real-world scenarios involving multiple individuals interacting simultaneously. This advancement holds significant potential for improving diagnostic accuracy, informing targeted interventions, and deepening our understanding of the mechanisms underlying impulsive motor behavior.

This study aimed to develop a cutting-edge automated methodology for assessing motor impulsivity in multi-person environments, involving the following key components:

- Developing and evaluating a computer vision-based framework that leverages the power of the YOLO object detector and ByteTrack multiobject tracking algorithms to accurately detect and track the movements of multiple individuals simultaneously from video data.
- Exploring the incorporation of contextual information, such as environmental factors and social interactions, to enhance the model's ability to recognize and differentiate impulsive motor behaviors in various real-world scenarios.
- Assessing the robustness and generalizability of the proposed approach by evaluating its performance across diverse movement quantification scenarios and subject characteristics, including varying levels of occlusion, interaction dynamics, and individual movement styles.
- Conducting a comprehensive comparative analysis against existing quantitative techniques, highlighting the potential advantages and limitations of our multi-person detection and tracking approach, particularly in its ability to facilitate accurate and objective assessments of motor impulsivity in complex, real-world scenarios involving multiple interacting individuals.

## 2. Methods

The process of assessing movements that characterize impulsive motor skills has great challenges; in addition to complexity and dynamism, there are also limitations to the approach described above related to the problem of complexity in terms of the number of subjects studied in real-world conditions. We proposed a framework that optimizes the use of deep learning for the analysis and classification of related movements. This section explains in detail each stage that we go through, both in terms of the technique and the algorithm that we use. Many things are involved, including data collection techniques, data pre-processing, feature extraction, post-processing, model training, classification, and model evaluation. Figure 1 illustrates the data generation process.

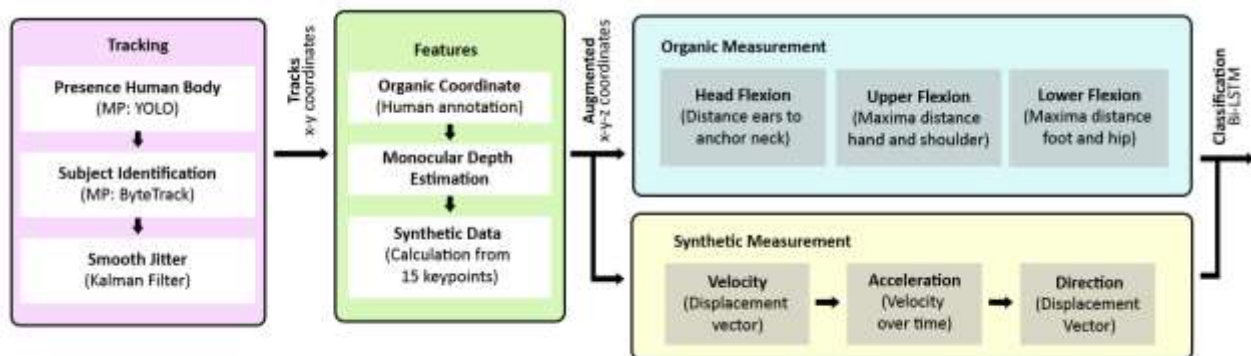


Figure 1. The process of generating data utilizes a step-by-step methodology. The video processing technique generates x,y, and z coordinates at each key point of the body, which are then separated into organic and synthetic components. The organic and synthetic components were processed independently and the resulting dataset was fed into the Bi-LSTM model.

### 2.1. Data Collection

This research used two datasets as data sources: one public dataset consisting of 84 videos obtained from YouTube. These videos show the behavior of one subject or several subjects, which can be categorized as having impulsive motor characteristics. One peculiarity of these videos is that they have a minimum number of subjects of two. The video was collected via the YouTube Data API using ten relevant keywords: “impulsive,” “fighting,” “raging,” “self-injury,” “hyperactivity,” “hyperkinetics,” “restlessness,” “group,”

“class,” and “playing.” The eight most suitable videos for each keyword were selected. The second dataset, as a complement, was a collection of 30 personal videos from three children who had a tendency to be motor impulsive with various symptoms. Each video included in the dataset has special criteria, namely that it has a minimum duration of 10s, and the behavioral symptoms of motor impulsivity must be clearly visible.

In total, we collected 135 video clips, and nearly 70% of the participants showed psychological disorders, exhibiting sudden behavior, unusual movements, or excessive activity. Each video was carefully annotated with metadata, such as location context, perceived environmental stimuli, and impulse duration. This comprehensive annotation process was carried out by the first author and a team consisting of six kindergarten teachers and six primary school teachers aged between 20 and 35 years. Each video was annotated by the first author and three teachers.

## 2.2. Preprocessing

To ensure a robust and diverse training dataset, thereby enhancing the generalization capabilities of the model, a strategic data augmentation process was implemented. Given the inherent challenges in acquiring a large-scale dataset of motoric impulsivity instances, particularly in multiperson scenarios, we employed a technique that effectively doubled the size of our initial dataset. Specifically, we reversed each video along the horizontal axis, mirroring the movements of the individuals, while preserving the corresponding labels associated with the level of motor impulsivity exhibited. Consequently, the augmented dataset comprising 270 videos provided a more comprehensive representation of motoric impulsivity patterns, encompassing a wider range of spatial configurations and movement variations.

## 2.3. Data Quantification

YOLO (You Only Look Once) is a cutting-edge, real-time object detection system that operates quickly. It utilizes a single neural network to simultaneously predict bounding boxes and directly classify objects from full images in a single pass. In our study, we employed the latest stable version, YOLOv8, to estimate the human poses in video footage. The model analyzes a video stream to detect and identify key landmarks on the human body in a 2D coordinate space with high accuracy and speed. Figure 2 depicts the numbering scheme for the body keypoints detected by YOLOv8, which follows the COCO (Common Objects in Context) standard.

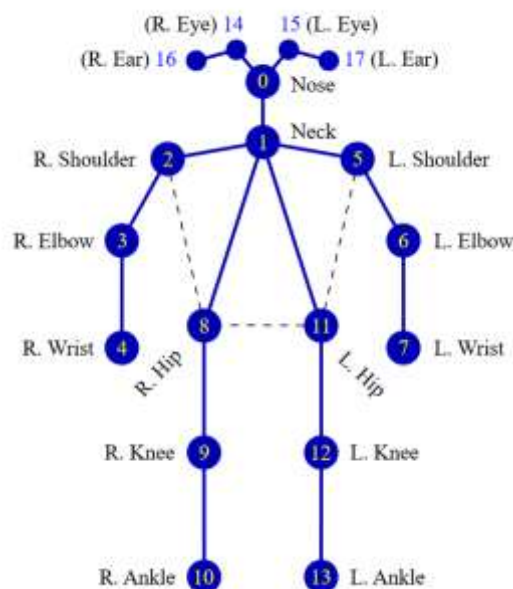


Figure 2. Key points for human poses according to COCO output format (R/L: right/left). [28]

This study concentrates on employing particular keypoints from YOLO to examine the movement organs, namely, the ears, knees, ankles, wrists, and elbows. This research also incorporated several anchor key points as reference points for relative movement. The keypoints used in this study are listed in table 2. Upon identifying the keypoints, the subsequent step involves implementing keypoint tracking in the video, providing labels for motion object tracking (MOT) identification, and mitigating the jitter that frequently arises during

the tracking process. The procedure is described in Algorithm 1.

Table 1. Utilized Keypoints

Type	#KP	Name
Peripheral	3	Right Elbow
	4	Right Wrist
	6	Left Elbow
	7	Left Wrist
	9	Right Knee
	10	Right Angkle
	12	Left Knee
	13	Left Ankle
	16	Right Ear
	17	Left Ear
Anchor	1	Neck
	2	Right Shoulder
	5	Left Shoulder
	8	Right Hip
	11	Left Hip

---

**ALGORITHM 1: KEYPOINT ANOTATION**

---

```

1  Open video capture with the provided video file path
2  While video capture is opened:
3      Read a frame from the video
4      If frame is successfully read:
5          Get the current timestamp in milliseconds
6          Run YOLOv8 tracking and ByteTrack on the frame with specified parameters
7          If keypoints are detected in the results:
8              Foreach keypoint in the results:
9                  Get the object ID if available, otherwise set to -1
10                 Flatten and convert keypoint to list
11                 Append timestamp, object ID, and keypoints to keypoints data
12             Visualize the results on the frame
13             Write the annotated frame to the output video
14             Display the annotated frame
15         Else:
16             Break the loop

```

---

A total of 270 videos from the preprocessing stage were input into YOLO. Figure 3 shows the interim results of this process.



Figure 3. The video clips processed using YOLO display an overlay of skeleton keypoints that are subsequently quantized.

When generating skeleton keypoints, YOLO frequently encounters difficulties due to interference with its environment, which can lead to inaccurate object identification and resulting jitter. To address this problem, we employed Kalman filters. [29]. It iteratively forecasts and determines the keypoint state using unreliable data by setting the initial keypoint location, predicting the state for each frame, and refining it by leveraging the detected keypoints. This method enables seamless keypoint tracking. Algorithm 2 is executed to handle jitter in the skeleton data.

---

**ALGORITHM 2: JITTER HANDLING**

---

```

1  Init a dictionary to separate keypoints by object ID
2  Foreach data in keypoints data:
3      Extract timestamp, object ID, and keypoints
4      If object ID is not in dictionary:
5          | Init an empty list for the object ID
6          | Append timestamp and keypoints to the object ID list
7  Init an empty list for smoothed keypoints data
8  Foreach object ID and its keypoints in the dictionary:
9      | Apply smoothing to the keypoints
10     | Foreach smoothed keypoint:
11     | | Append object ID and smoothed keypoint to smoothed keypoints data

```

---

We evaluated each video according to the following requirements to assess the movement aspects:

- (1) Head Flexion: Local minimum distance between the ears and neck keypoints.
- (2) Upper Bending: Local maxima of the distance between the keypoints of the arm and shoulder.
- (3) Lower Bending: Local maxima of the distance between the foot and hip keypoints.
- (4) The ground truth was manually set using frame-by-frame inspection.

The model records the keypoints and exports data as comma-separated values (.csv) files. We split the data into two types, based on the generation process. First, organic data containing the position of each keypoint per unit time were presented as two-dimensional [x,y] coordinates for the peripheral and anchor keypoints. We also obtained the [z] coordinate by leveraging the depth of the keypoint using Monocular Depth Estimation. Second, synthetic data were obtained by calculating the velocity, acceleration, motion trajectory, flexion distance, and flexion abduction from organic keypoint data.

#### 2.4. Data Analysis and Evaluation

- 1) *Monocular Depth Estimation*: To assess the distance of objects from a camera using a single image, we employed a method known as Monocular Depth Sensing (MiDaS)[30]. This process entails the utilization of a neural network that has been extensively trained on datasets of images that possess corresponding depth maps. Upon presenting a new image, the network could predict the depth of each pixel with high accuracy. In our case, we utilized the MiDaS model, which is a pre-trained neural network capable of estimating the depth from a single image. Specifically, we employ the *DPT\_Large* variant of the model to increase precision. Every frame in the video is input into the depth model to obtain a depth map. The depth map is a 2D array that represents the estimated distance from the camera for each pixel in the frame. By referencing the corresponding depth value for each detected keypoint (x, y), we can determine the z-coordinates of that keypoint. Algorithm 3 illustrates the study of MiDaS.

---

**ALGORITHM 3: MONOCULAR DEPTH ESTIMATION**

---

```

1  input_batch = APPLY depth_transforms TO frame
2  MOVE input_batch TO device (GPU/CPU)
3  WITH torch.no_grad():
4      | prediction = depth_model(input_batch)
5      | depth_map = RESIZE prediction USING bicubic interpolation
6      | REMOVE extra dimension from depth_map
7  depth_map = CONVERT depth_map TO numpy array

```

---

- 2) *Velocity Calculation:* Velocity serves as a crucial factor in distinguishing between impulsive and controlled movements. Typically, higher velocities are suggestive of impulsive behavior. To determine the velocity of a keypoint's movement, it is necessary to calculate the distance between its initial and final positions, and the amount of time it takes to move from one position to another. It is essential to note that the initial position of the keypoint is  $(x_1, y_1, z_1)$  and the final position is  $(x_2, y_2, z_2)$ , the transposition on each axis can be calculated using  $\Delta x = x_2 - x_1$  on the x axis,  $\Delta y = y_2 - y_1$  on the y axis,  $\Delta z = z_2 - z$  on the z axis. The displacement distance was calculated using the Pythagorean equation, which determines the displacement length.

$$d = \sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2} \quad (1)$$

The following formula can be used to determine the velocity:

$$v = \frac{d}{t} = \frac{\sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2}}{t} \quad (2)$$

To represent the subject's motion more accurately, the standard deviation (*std*) of the velocities was calculated. This could provide an insight into the variability of motion. A higher standard deviation may indicate more erratic or irregular movement patterns.

$$std = \sqrt{\frac{1}{N} \sum_{i=1}^N (v_i - \bar{v})^2} \quad (3)$$

Where N is the total number of samples,  $v_i$  is the velocity of the sample  $i$ , and  $\bar{v}$  is the mean velocity. This formula calculates the square root of the average of the squared differences between each velocity value and the mean velocity. It measures the dispersion or spread of velocity values around the mean velocity.

- 3) *Acceleration Calculation:* Acceleration of a keypoint's movement in a three-dimensional space can be determined by calculating the change in vector velocity over a specific time frame. Acceleration monitoring can reveal sudden and jerky movements that are often indicative of impulsive behavior. To determine the initial and final velocities of the keypoint movement before the calculation, the change in the velocity per unit time was measured. It is assumed that the initial velocity of the keypoint is  $v_1 = v_{1x}\hat{i} + v_{1y}\hat{j} + v_{1z}\hat{k}$  and the final velocity is  $v_2 = v_{2x}\hat{i} + v_{2y}\hat{j} + v_{2z}\hat{k}$  with  $v_{1x}, v_{1y}, v_{1z}, v_{2x}, v_{2y}, v_{2z}$  are the components of the velocity along each axis. The change in the velocity was calculated by subtracting the initial velocity from the final velocity for each component. On the x-axis, the formula applied is  $\Delta v_x = v_{2x} - v_{1x}$  whereas on the y-axis the formula applied is  $\Delta v_y = v_{2y} - v_{1y}$ , and on the z-axis is  $\Delta v_z = v_{2z} - v_{1z}$ . The acceleration of motion is the change in the total velocity divided by the total time. This was calculated using the following equation:

$$a = \frac{\Delta v_x}{\Delta t} \hat{i} + \frac{\Delta v_y}{\Delta t} \hat{j} + \frac{\Delta v_z}{\Delta t} \hat{k} \quad (4)$$

We then used the data to calculate the jerk data. Jerk data is the rate of change of acceleration and can be calculated as the derivative of acceleration with respect to time. Analyzing jerk can provide insights into the smoothness or abruptness of motion transitions. We divide three metrics of jerk: a) *Jerk Mean*: This average rate of change of acceleration over time for a specific keypoint. It provides an indication of the average smoothness or abruptness of motion changes at that keypoint; b) *Jerk Max*: This represents the maximum rate of change of acceleration observed over time for a specific keypoint. It indicates the maximum abruptness or intensity of motion changes at that keypoint; c) *Jerk Min*: This represents the minimum rate of change of acceleration observed over time for a specific keypoint. This indicates that the minimum abruptness or intensity of motion changes at that key point.

$$Jerk\ Mean = \frac{1}{n} \sum_{i=1}^n \frac{d^3x}{dt^3i} \quad (5)$$

$$Jerk\ Min = \min\left(\frac{d^3x}{dt^31}, \frac{d^3x}{dt^32}, \dots, \frac{d^3x}{dt^3n}\right) \quad (6)$$

$$Jerk\ Max = \max\left(\frac{d^3x}{dt^31}, \frac{d^3x}{dt^32}, \dots, \frac{d^3x}{dt^3n}\right) \quad (7)$$

Where  $\frac{d^3x}{dt^3_i}$  is the jerk at time  $i$ , and  $n$  is the total number of samples. These statistics help quantify how smoothly or abruptly a keypoint's motion changes over time. They provide insights into the dynamics of motion and can be useful for various applications, such as motion analysis, gesture recognition, and activity monitoring.

- 4) *Direction Calculation:* Calculating movement direction is crucial, as it provides valuable information about the direction and orientation of movements, which can indicate specific impulsive behaviors or patterns. To ascertain the direction of the displacement, it is necessary to compute the displacement vector and confirm its direction. The displacement vector, which connects the initial position of the keypoint to its final position, is calculated using the following formula:

$$\vec{A} = \Delta x \hat{i} + \Delta y \hat{j} + \Delta z \hat{k} \quad (8)$$

The unit vectors of the x, y, and z axes are denoted by  $\hat{i}, \hat{j}, \hat{k}$ . The direction of the displacement vector can be determined using its magnitude and dot product. To find the angle between the vector and the x-, y-, and z-axes, the arccosine of the dot product is divided by the vector size.

$$\theta_x = \cos^{-1} \frac{\Delta x}{\sqrt{x^2 + y^2 + z^2}} \quad (6)$$

$$\theta_y = \cos^{-1} \frac{\Delta y}{\sqrt{x^2 + y^2 + z^2}} \quad (7)$$

$$\theta_z = \cos^{-1} \frac{\Delta z}{\sqrt{x^2 + y^2 + z^2}} \quad (8)$$

- 5). *Data Analysis:* Our model was implemented by allocating the processed dataset to three segments: 60% for training, 20% for validation, and 20% for testing. Table 2 lists the hyperparameter configurations used in building the Bi-LSTM architecture and the optimal settings determined through performance evaluation based on the validation dataset. These hyperparameters were used to train the model with the best performance during the experiment.

Table 2. Bi-LSTM Hyperparameter Configuration

Hyperparameter	Value
Activation function	Sigmoid
Loss function	Binary Crossentropy
Optimizer	ADAM
Learning rate	0.01
Epsilon	1.e-07
learning rate decay	0.01
Epochs	10, 30, 50
Dropout	0.5
Batch Size	64
Nodes per Layer	100

A confusion matrix was used to assess the performance of the algorithm by comparing its predicted and actual class instances. The values of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) were determined. In our research, we employed various metrics, including Accuracy, Sensitivity (recall), Precision, and Area under the ROC Curve (AUC), to evaluate the algorithm. This metric is calculated as follows:

$$Accuracy = \frac{(TP + TN)}{(TP + FP + TN + FN)} \quad (9)$$

$$Precision = \frac{(TP)}{(TP + FP)} \quad (10)$$



$$Recall = \frac{(TP)}{(TP + FN)} \quad (11)$$

### 3. Result and Analyses

Our framework offers notable improvements over existing techniques for assessing and categorizing motoric impulsivity in a multi-person environment. To evaluate the efficacy of our approach, we employed a comprehensive dataset comprising videos from diverse sources such as online platforms and specially recorded sessions. We carefully curated this dataset to capture a broad range of impulsive movements exhibited by the children in various settings. By doing so, we ensured a comprehensive evaluation of the performance of our model.

#### 3.1. Movement Detection and Quantification Results

- 1) *Head Flexion*: Head flexion refers to the act of bending the head towards the left or right shoulder, which is often observed in impulsive behaviors such as fidgeting or restlessness. This movement could serve as a reliable indicator of motor impulsivity. To accurately measure head lateral flexion, frame-by-frame video analysis was performed with specific ear landmarks tracked to quantify the degree and frequency of the movement. Excessive or repetitive head lateral flexion can lead to a disruption in focus and interfere with daily activities and may even be indicative of underlying conditions such as ADHD or anxiety disorders. The quantified results for head flexion in the first five private videos are presented in Table 3.

Table 3. The Movement Quantification of Head Flexion

Video	n	$\bar{d}$	J Max	J	std
#1	2	0.09	140.37	22.83	144.05
		0.05	139.92	24.56	148.13
#2	2	0.19	206.77	70.11	169.06
		0.17	105.14	15.14	109.62
#3	4	0.17	146.37	27.64	151.68
		0.16	143.77	35.89	162.81
		0.14	127.69	34.69	165.26
		0.09	745.90	76.70	239.07
		0.25	126.44	15.05	130.27
#4	6	0.15	654.04	27.91	159.46
		0.14	190.67	29.11	165.66
		0.11	182.33	32.77	186.09
		0.08	140.09	41.18	201.51
		0.10	428.93	72.36	210.80
		0.30	478.98	30.28	172.16
#5	6	0.18	439.27	40.81	203.38
		0.13	542.24	26.61	153.73
		0.11	205.67	20.34	122.83
		0.36	65.80	9.00	80.92
		0.19	139.01	28.54	195.33

- 2) *Upper Bending*: In addition to monitoring head movements, we closely observed the motion of the brachium (upper arm) and the antebrachium (forearm). Excessive or abnormal movements in these regions may indicate impulsive behaviors such as repetitive or self-stimulatory behaviors. To gauge the extent and range of these movements, we carefully tracked the angles of the elbow and shoulder joints and established neutral zero positions. Monitoring the brachium and antebrachium is crucial because impulsive behaviors may manifest not only in the wrist but also in a broader range of motion involving the entire arm. This information is critical for recognizing potential impulsive behaviors connected to arm movements that can hinder daily activities, social interactions, or educational settings. Table 4 presents the results of the quantification of the movements in the upper bending section.

Table 4. The Movement Quantification of Upper Bending

video	n	Brachium L				Brachium R				Antebrachium L				Antebrachium R			
		$\bar{d}$	$J_{max}$	$\bar{J}$	$std$	$\bar{d}$	$J_{max}$	$\bar{J}$	$std$	$\bar{d}$	$J_{max}$	$\bar{J}$	$std$	$\bar{d}$	$J_{max}$	$\bar{J}$	$std$
#1	2	0.32	126.89	18.33	118.17	0.38	40.32	5.67	35.72	0.32	135.43	20.05	128.81	0.38	113.14	5.96	52.60
		0.01	201.66	7.93	75.63	0.11	201.12	26.31	119.95	0.04	205.81	25.62	154.31	0.41	200.81	19.25	103.93
#2	2	0.14	202.95	67.70	203.91	0.12	219.37	70.64	199.26	0.09	111.02	46.77	190.72	0.13	216.07	51.04	200.70
		0.32	109.12	15.46	106.81	0.35	208.78	20.97	129.65	0.27	111.07	18.19	124.36	0.39	207.53	0.77	137.21
#3	4	0.37	112.91	21.19	106.64	0.43	134.87	27.96	129.63	0.37	121.94	21.52	113.09	0.44	126.55	28.55	134.52
		0.39	116.55	24.89	129.45	0.40	143.58	31.60	143.55	0.38	125.96	27.99	140.29	0.41	141.40	34.30	143.39
		0.31	152.90	35.99	168.74	0.32	200.10	35.78	163.11	0.25	163.57	39.99	173.13	0.29	134.16	38.04	161.77
		0.25	795.70	72.51	217.92	0.20	196.33	48.65	170.86	0.21	732.36	76.70	237.50	0.18	211.16	47.98	182.49
#4	6	0.31	0	0	0	0.42	828.61	37.57	179.63	0.31	123.73	30.64	159.95	0.37	532.08	30.96	153.43
		0.40	0	0	0	0.51	853.86	38.92	179.17	0.39	380.61	32.96	159.24	0.46	139.47	35.35	172.82
		0.29	0	0	0	0.31	145.06	40.72	188.45	0.29	153.12	27.62	164.09	0.30	250.17	41.96	184.79
		0.31	149.85	26.93	158.27	0.28	222.82	33.36	180.35	0.31	153.26	35.31	173.61	0.31	241.69	36.20	186.56
		0.11	153.60	38.46	204.35	0.08	444.14	49.73	197.15	0.1080	131.46	27.17	168.27	0.07	481.17	41.11	174.83
		0.08	478.51	56.50	191.78	0.05	148.64	38.27	186.96	0.09	132.46	41.20	187.95	0.04	141.19	38.40	187.92
#5	6	0.43	186.15	17.51	97.00	0.54	169.36	18.32	102.29	0.45	155.20	18.62	104.21	0.53	143.64	18.58	104.47
		0.49	242.43	27.05	139.60	0.63	213.41	26.99	143.11	0.52	179.82	27.32	144.97	0.61	163.81	27.03	145.50
		0.61	170.51	20.85	111.79	0.80	200.28	21.73	114.81	0.64	171.33	21.71	115.53	0.80	190.09	21.86	115.49
		0.26	173.10	25.90	140.92	0.43	203.48	17.08	89.79	0.30	181.99	24.83	136.17	0.45	194.95	16.86	87.87
		0.44	123.35	15.67	123.95	0.57	62.41	8.06	68.45	0.43	128.36	18.34	133.96	0.56	67.68	7.58	65.70
		0.31	125.47	13.28	89.20	0.37	206.78	12.56	73.52	0.21	130.62	11.04	82.62	0.40	212.18	15.23	82.72

Table 5. The Movement Quantification of Lower Bending

video	n	Thigh L				Thigh R				Calf L				Calf R			
		$\bar{d}$	$J_{max}$	$\bar{J}$	$std$	$\bar{d}$	$J_{max}$	$\bar{J}$	$std$	$\bar{d}$	$J_{max}$	$\bar{J}$	$std$	$\bar{d}$	$J_{max}$	$\bar{J}$	$std$
#1	2	0.56	63.59	8.52	50.24	0.58	43.14	7.07	42.77	0.64	70.28	10.21	58.81	0.66	54.83	8.42	51.87
		0.04	145.60	8.73	145.01	0.04	234.46	11.75	143.57	0.038	150.51	9.41	161.08	0.04	253.58	12.60	157.02
#2	2	0.31	209.19	83.07	181.74	0.27	217.48	77.97	186.63	0.23	622.32	120.48	235.87	0.15	226.62	99.27	211.27
		0.48	131.39	11.38	86.71	0.47	130.50	14.69	109.14	0.45	199.06	25.23	151.23	0.47	135.45	19.39	131.33
#3	4	0.61	154.02	22.65	120.09	0.62	147.24	22.24	108.03	0.63	295.49	27.02	140.99	0.66	0	0	0
		0.62	149.89	22.74	125.18	0.62	143.86	23.95	116.47	0.64	0	0	0	0.66	0	0	0
		0.46	235.54	34.73	167.26	0.47	225.22	34.39	163.12	0.44	242.88	40.28	196.24	0.46	0	0	0
		0.28	298.56	78.78	210.97	0.30	293.48	73.02	201.47	0.18	495.63	66.70	222.01	0.19	276.09	77.02	216.19
#4	6	0.19	145.02	22.53	146.58	0.25	147.95	33.84	178.69	0.06	148.87	20.26	144.04	0.08	147.56	23.19	155.47
		0.31	122.89	28.89	161.99	0.38	138.30	36.44	176.81	0.13	151.72	28.89	174.73	0.16	129.36	24.08	162.25
		0.27	166.82	26.93	161.80	0.28	240.84	32.19	171.04	0.16	128.77	29.33	176.80	0.16	182.15	34.93	181.28
		0.33	165.49	34.05	162.96	0.35	123.85	31.63	156.31	0.20	130.87	37.23	180.22	0.17	189.14	37.29	182.79
		0.13	143.79	31.02	172.66	0.13	146.85	33.51	177.60	0.21	148.29	41.62	205.04	0.19	148.82	38.91	200.26
		0.09	148.26	38.01	198.18	0.09	145.80	37.63	198.25	0.11	150.25	46.40	221.22	0.10	339.72	51.96	220.44
#5	6	0.62	142.23	17.85	101.45	0.69	149.30	19.25	108.83	0.72	142.33	16.43	93.23	0.77	177.02	19.75	111.00
		0.67	149.48	25.90	140.00	0.75	168.32	27.96	149.25	0.77	164.40	24.12	128.27	0.84	218.49	29.26	153.67
		0.79	179.10	20.31	110.07	0.91	223.89	22.30	118.62	0.89	164.01	18.62	100.84	1.01	253.90	22.98	122.27
		0.53	160.86	16.67	89.09	0.58	225.24	16.87	89.87	0.63	140.37	16.59	90.26	0.68	254.05	16.40	86.75
		0.66	60.83	8.38	71.26	0.69	55.29	7.88	66.71	0.74	62.69	8.35	71.02	0.76	53.22	7.40	62.28
		0.39	119.14	22.99	156.95	0.41	109.86	21.34	144.97	0.47	125.93	23.32	163.34	0.48	116.12	21.67	151.58

3) *Lower Bending*: The collective term for the movements of the thigh and calf, such as flexion and abduction, that occur during activities, such as walking and running, is referred to as lower bending. It is essential to measure these movements because they can provide valuable insights into gait patterns, coordination issues, and excessive movements related to impulsivity. By tracking anatomical landmarks and calculating joint angles over time, the range of motion and coordination of thigh, calf, and foot movements can be evaluated. The quantification results of lower bending movements involving the thigh and calf are presented in Table 5.

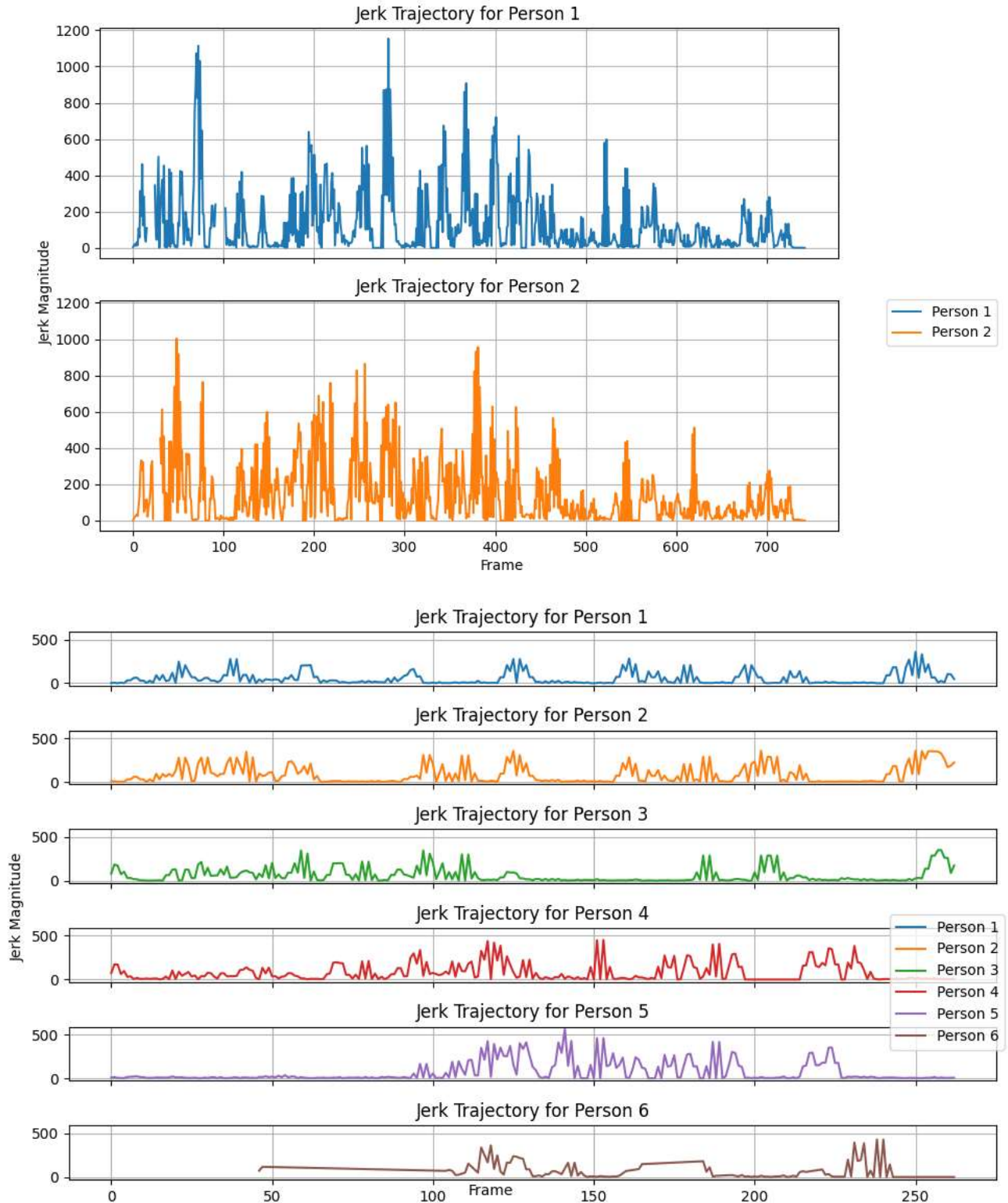


Figure 4. (a) Jerk trajectory for two subjects in the video; (b) Jerk trajectory for six subjects in the video.

In our investigation of motion dynamics, we transformed raw data into a captivating visual narrative utilizing 15 videos that exhibited diverse numerical traits. As shown in Figure 4, we carefully selected two distinct samples for visual representation. By employing a 2D line plot, we displayed these contrasting individual side by side, providing a comprehensive overview of their disparate trajectories. This visual depiction serves as a powerful portrayal of their dual nature, accentuating the dichotomy between motor impulsiveness and non-impulsiveness. Delving deeper into the complexities, we examined the intricate relationship between acceleration changes, vividly captured by fluctuations in jerk magnitude over time. Each data point on the graph reveals a story of motion outlining the cadence of smooth transitions and sudden shifts. Through this visual journey, we uncovered the hidden layers of movement, illuminating the essence of jerk as a discerning metric for characterizing the nuances of motion dynamics.

### 3.2. Performance

An extensive evaluation of the proposed model was conducted using the quantitative data presented earlier and employing the optimal configuration detailed in Table 3. The performance of the system on the test dataset was prominent, achieving an accuracy of up to 95.43%. This figure shows the efficacy of the model in classifying the sample and determining the proportion of correct predictions in relation to the amount of observed data, as shown in Figure 5.

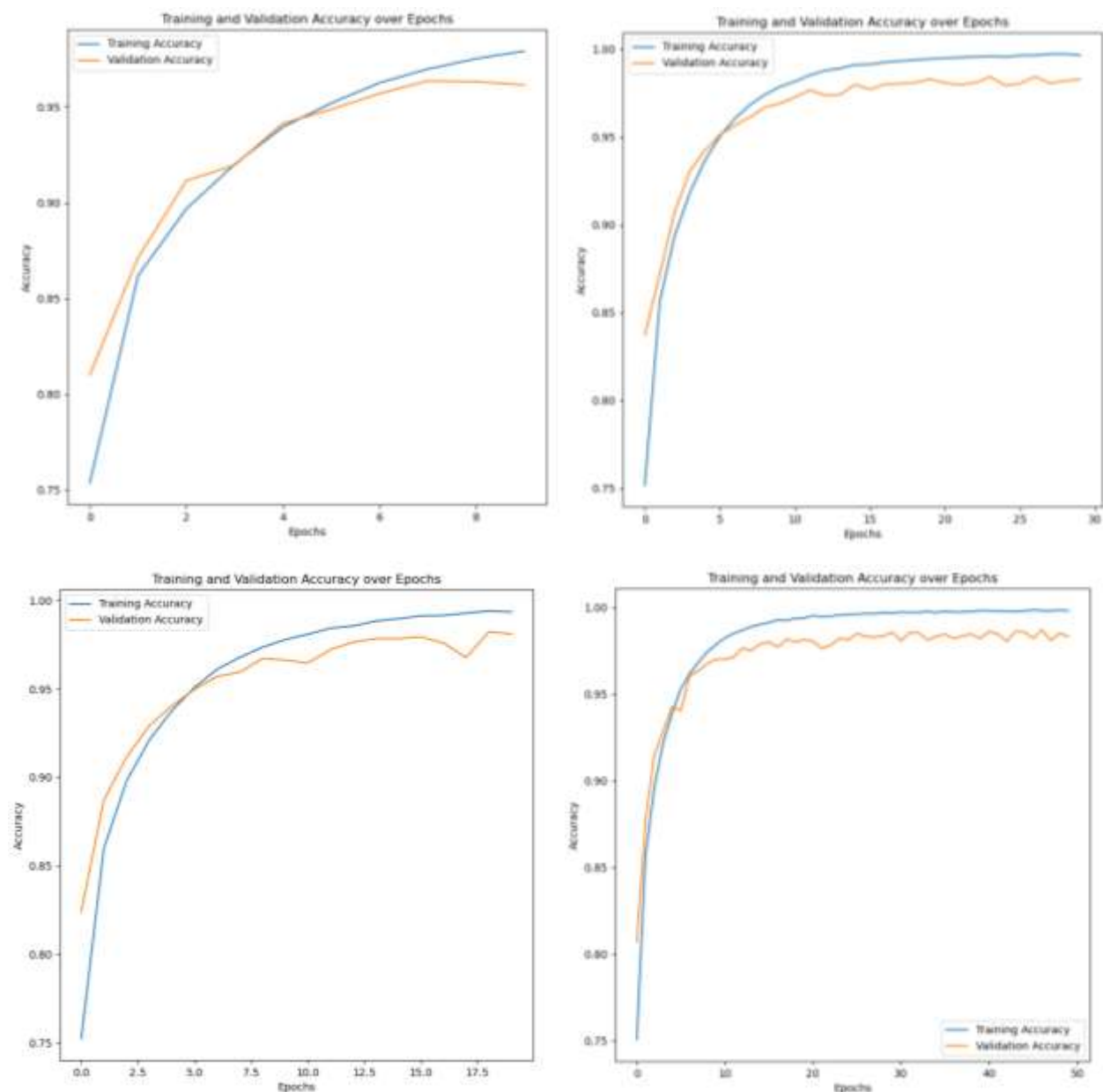


Figure 5 Accuracy curves between training and validation for 10, 20, 30, and 50-epochs

The performance of our classification model is demonstrated by an AUC value of 0.96, as shown in Figure 6. This value provides evidence that our model outperforms chance, as indicated by the dashed diagonal line in the curve.

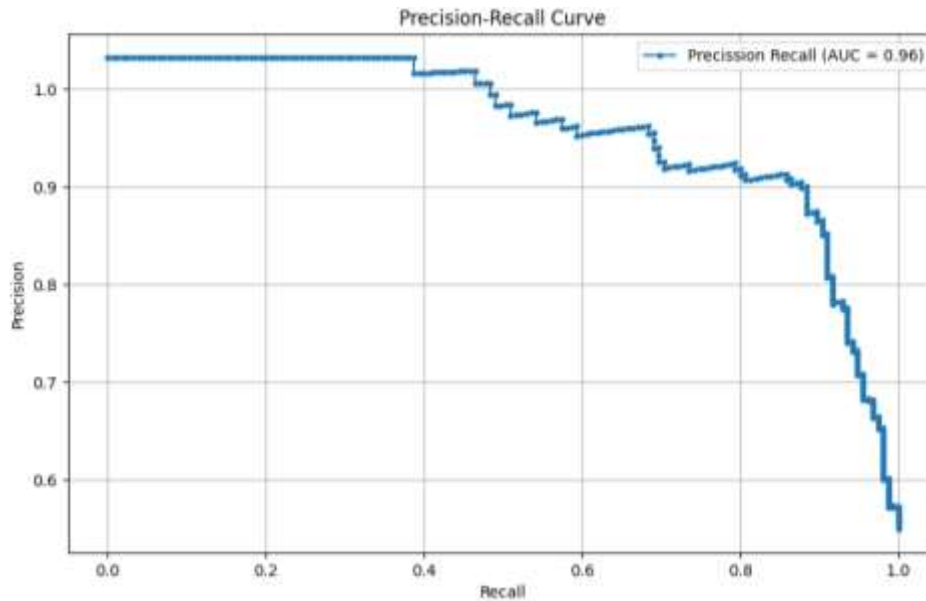


Figure 6. Precision-recall curves for training and testing during 50-epochs.

The confusion matrix revealed that the model rarely misidentifies non-impulsive movements as impulsive, and effectively captures impulsive movements. This demonstrates its ability to distinguish between genuine impulsive and fast-regular movements. Taken together, the performance metrics, including accuracy, precision, recall, and F1 scores, indicate a robust, high-performance model for accurately categorizing instances across various classes.

In the proposed work, significant progress was made compared with similar architectures, as shown in Table 6. This table effectively compares the proposed design with those presented in [31] and [32].

Table 6. Accuracy Comparison with Previous Study

Reference	Method Used	Description of the Experiment	Accuracy
[31]	YOLOv7, BoT-SORT	combination of YOLO with ByteTrack-SORT for pedestrian multi-object tracking, showcasing the integration of these two algorithms for improved tracking accuracy.	79.11%
[32]	MultiFusedNet (CNN, K-Means), Bi-LSTM	The experiments aimed to perform individual behavior analysis in situations where lighting changes are diverse and groups of students are intermingled.	96.67%
Our Research	YoloV8, ByteTrack, Bi-LSTM	Assesing motor impulsivity of multiple children in same stage using computer vision and deep learning. The model, incorporating YoloV8 and a Bi-LSTM architecture. This highlights the effectiveness of combining these technologies for precise motor impulsivity analysis, valuable for enhancing safety measures.	95.43%

Our research encounters certain limitations that require additional exploration. One limitation is that accurately interpreting motor impulsivity necessitates consideration of the environmental context in which activity or movement takes place. Relying on predefined definitions may not offer an accurate assessment of observed movements in a specific context. To enhance the generalizability and robustness of our method, future studies should concentrate on incorporating environmental context information into the analysis

pipeline, such as scene understanding or situational awareness modeling. This would offer a more comprehensive understanding of the contextual factors that affect the observed behaviors.

Furthermore, investigating the integration of multimodal data sources such as physiological signals or environmental factors could provide a more holistic understanding of motor impulsivity and its triggers. This may lead to personalized and effective interventions.

#### 4. Conclusion

Our study successfully demonstrated the use of advanced computer vision and deep learning techniques to accurately identify and quantify motor impulsivity in complex, multi-person environments. By integrating state-of-the-art object detection (YOLOv8) and multi-object tracking (ByteTrack) algorithms, we developed a robust framework capable of reliably tracking and analyzing the movements of multiple individuals simultaneously.

The incorporation of the Bi-LSTM architecture further enhanced our model's ability to capture and classify intricate patterns of motor impulsivity, resulting in an accuracy rate of 95.43% in distinguishing impulsive motor behaviors from typical movements. This performance was facilitated by the capacity of our model to extract and analyze a comprehensive set of kinematic features, including abrupt changes in body position, erratic velocity/acceleration profiles, and recurring and disruptive motion patterns.

Our approach addresses a long-standing limitation in the field by enabling objective and precise assessments of motor impulsivity in naturalistic, multi-person scenarios, transcending the constraints of traditional subjective methodologies. These findings pave the way for transformative applications across diverse domains such as designing inclusive and accommodating learning environments, implementing proactive safety measures in recreational spaces, and fostering supportive home settings that nurture children's well-being and development.

Moreover, our work heralds a paradigm shift in pediatric psychology and developmental neuroscience, ushering in a new era of data-driven precision interventions tailored to individual needs. By precisely quantifying and characterizing motor impulsivity, our approach enables the development of personalized therapeutic strategies and targeted interventions for related disorders such as Attention-Deficit/Hyperactivity Disorder (ADHD).

While our study yielded excellent results, we acknowledge the inherent limitations of our work, including the relatively modest dataset size and focus on a specific age group. Future research endeavors should aim to expand the scope of our approach by exploring larger and more diverse datasets encompassing a wider range of age groups, cultural backgrounds, and neurodevelopmental conditions. Additionally, investigating the generalizability of our framework to other domains, such as geriatric care and sports performance analysis, could yield invaluable insights and applications.

Our research not only addresses a critical need in the field but also serves as a catalyst for further innovation and multidisciplinary collaboration, paving the way for a future in which cutting-edge technology seamlessly integrates with human-centric interventions, fostering a deeper understanding of neurodiversity and promoting inclusive, supportive environments for individuals of all abilities.

#### References

- [1] E. K. Farran *et al.*, "Is the Motor Impairment in Attention Deficit Hyperactivity Disorder (ADHD) a Co-Occurring Deficit or a Phenotypic Characteristic?," *Adv Neurodev Disord*, vol. 4, no. 3, pp. 253–270, Sep. 2020, doi: 10.1007/s41252-020-00159-6.
- [2] O. Grimm *et al.*, "Impulsivity and Venturesomeness in an Adult ADHD Sample: Relation to Personality, Comorbidity, and Polygenic Risk," *Front Psychiatry*, vol. 11, Dec. 2020, doi: 10.3389/fpsy.2020.557160.
- [3] E. K. Edmiston *et al.*, "Assessing Relationships Among Impulsive Sensation Seeking, Reward Circuitry Activity, and Risk for Psychopathology: A Functional Magnetic Resonance Imaging Replication and Extension Study," *Biol Psychiatry Cogn Neurosci Neuroimaging*, vol. 5, no. 7, pp. 660–668, Jul. 2020, doi: 10.1016/j.bpsc.2019.10.012.
- [4] X. Chen and S. Li, "Serial mediation of the relationship between impulsivity and suicidal ideation by depression and hopelessness in depressed patients," *BMC Public Health*, vol. 23, no. 1, p. 1457, Jul. 2023, doi: 10.1186/s12889-023-16378-0.

- [5] J. Salles *et al.*, “Indirect effect of impulsivity on suicide risk through self-esteem and depressive symptoms in a population with treatment-resistant depression: A FACE-DR study,” *J Affect Disord*, vol. 347, pp. 306–313, Feb. 2024, doi: 10.1016/j.jad.2023.11.063.
- [6] T. Veliki, Z. Užarević, and S. Dubovicki, “Self-Evaluated ADHD Symptoms as Risk Adaptation Factors in Elementary School Children,” *Drustvena istrazivanja*, vol. 28, no. 3, pp. 503–522, Oct. 2019, doi: 10.5559/di.28.3.07.
- [7] A. Bandyopadhyay *et al.*, “Behavioural difficulties in early childhood and risk of adolescent injury,” *Arch Dis Child*, vol. 105, no. 3, pp. 282–287, Mar. 2020, doi: 10.1136/archdischild-2019-317271.
- [8] C. K. Conners, J. Pitkanen, and S. R. Rzepa, “Conners 3rd Edition (Conners 3; Conners 2008),” in *Encyclopedia of Clinical Neuropsychology*, New York, NY: Springer New York, 2011, pp. 675–678. doi: 10.1007/978-0-387-79948-3\_1534.
- [9] G. J. DuPaul, T. J. Power, A. D. Anastopoulos, and R. C. Reid, “Adhd Rating Scale-IV: Checklists, Norms, and Clinical Interpretation,” 1998. [Online]. Available: <https://api.semanticscholar.org/CorpusID:141673166>
- [10] G. J. DuPaul and G. Stoner, *ADHD in the schools: Assessment and intervention strategies, 3rd ed.* New York, NY, US: The Guilford Press, 2014.
- [11] G. Mirabella, C. Mancini, S. Pacifici, D. Guerrini, and A. Terrinoni, “Enhanced reactive inhibition in adolescents with non-suicidal self-injury disorder,” *Dev Med Child Neurol*, vol. 66, no. 5, pp. 654–666, May 2024, doi: 10.1111/dmcn.15794.
- [12] J. Bertram *et al.*, “Accuracy and repeatability of the Microsoft Azure Kinect for clinical measurement of motor function,” *PLoS One*, vol. 18, no. 1, p. e0279697, Jan. 2023, doi: 10.1371/journal.pone.0279697.
- [13] A. Jalal, M. A. K. Quaid, S. B. ud din Tahir, and K. Kim, “A Study of Accelerometer and Gyroscope Measurements in Physical Life-Log Activities Detection Systems,” *Sensors*, vol. 20, no. 22, p. 6670, Nov. 2020, doi: 10.3390/s20226670.
- [14] D. Kobsar *et al.*, “Wearable Inertial Sensors for Gait Analysis in Adults with Osteoarthritis—A Scoping Review,” *Sensors*, vol. 20, no. 24, p. 7143, Dec. 2020, doi: 10.3390/s20247143.
- [15] Y. Zhang *et al.*, “ByteTrack: Multi-object Tracking by Associating Every Detection Box,” 2022, pp. 1–21. doi: 10.1007/978-3-031-20047-2\_1.
- [16] D. Zhao, G. Su, G. Cheng, P. Wang, W. Chen, and Y. Yang, “Research on real-time perception method of key targets in the comprehensive excavation working face of coal mine,” *Meas Sci Technol*, vol. 35, no. 1, p. 015410, Jan. 2024, doi: 10.1088/1361-6501/ad060e.
- [17] A. Graves and J. Schmidhuber, “Framewise phoneme classification with bidirectional LSTM and other neural network architectures,” *Neural Networks*, vol. 18, no. 5–6, pp. 602–610, Jul. 2005, doi: 10.1016/j.neunet.2005.06.042.
- [18] A. Dey, S. Biswas, and D.-N. Le, “Recognition of Wh-Question Sign Gestures in Video Streams using an Attention Driven C3D-BiLSTM Network,” *Procedia Comput Sci*, vol. 235, pp. 2920–2931, 2024, doi: 10.1016/j.procs.2024.04.276.
- [19] K. Nguyen-Trong, H. N. Vu, N. N. Trung, and C. Pham, “Gesture Recognition Using Wearable Sensors With Bi-Long Short-Term Memory Convolutional Neural Networks,” *IEEE Sens J*, vol. 21, no. 13, pp. 15065–15079, Jul. 2021, doi: 10.1109/JSEN.2021.3074642.
- [20] R. P. Singh and L. D. Singh, “Dyhand: dynamic hand gesture recognition using BiLSTM and soft attention methods,” *Vis Comput*, Mar. 2024, doi: 10.1007/s00371-024-03307-4.
- [21] J. Wu, P. Ren, B. Song, R. Zhang, C. Zhao, and X. Zhang, “Data glove-based gesture recognition using CNN-BiLSTM model with attention mechanism,” *PLoS One*, vol. 18, no. 11, p. e0294174, Nov. 2023, doi: 10.1371/journal.pone.0294174.
- [22] X. Wu and Q. Ji, “TBRNet: Two-Stream BiLSTM Residual Network for Video Action Recognition,” *Algorithms*, vol. 13, no. 7, p. 169, Jul. 2020, doi: 10.3390/a13070169.
- [23] A.-A. Liu, Z. Shao, Y. Wong, J. Li, Y.-T. Su, and M. Kankanhalli, “LSTM-based multi-label video event detection,” *Multimed Tools Appl*, vol. 78, no. 1, pp. 677–695, Jan. 2019, doi: 10.1007/s11042-017-5532-x.
- [24] X. Tong, X. Tan, and X. Sun, “Abnormal behavior detection based on GCN-BiLSTM,” in *Third International Conference on Machine Learning and Computer Application (ICMLCA 2022)*, F. Zhou and S. Ba, Eds., SPIE, May 2023, p. 58. doi: 10.1117/12.2675168.
- [25] M. A. Soeleman, C. Supriyanto, D. P. Prabowo, and P. N. Andono, “Video Violence Detection Using LSTM and Transformer Networks Through Grid Search-Based Hyperparameters Optimization,”



- International Journal of Safety and Security Engineering*, vol. 12, no. 05, pp. 615–622, Nov. 2022, doi: 10.18280/ijss.120510.
- [26] F. Carrara, P. Elias, J. Sedmidubsky, and P. Zezula, “LSTM-based real-time action detection and prediction in human motion streams,” *Multimed Tools Appl*, vol. 78, no. 19, pp. 27309–27331, Oct. 2019, doi: 10.1007/s11042-019-07827-3.
- [27] W. Ullah, A. Ullah, I. U. Haq, K. Muhammad, M. Sajjad, and S. W. Baik, “CNN features with bi-directional LSTM for real-time anomaly detection in surveillance networks,” *Multimed Tools Appl*, vol. 80, no. 11, pp. 16979–16995, May 2021, doi: 10.1007/s11042-020-09406-3.
- [28] F.-C. Lin, H.-H. Ngo, C.-R. Dow, K.-H. Lam, and H. L. Le, “Student Behavior Recognition System for the Classroom Environment Based on Skeleton Pose Estimation and Person Detection,” *Sensors*, vol. 21, no. 16, p. 5314, Aug. 2021, doi: 10.3390/s21165314.
- [29] F. F. Dalimarta, Z. A. Hasibuan, P. N. Andono, Pujiono, and M. A. Soeleman, “Lower Body Detection and Tracking with AlphaPose and Kalman Filters,” in *Proceedings - 2021 International Seminar on Application for Technology of Information and Communication: IT Opportunities and Creativities for Digital Innovation and Communication within Global Pandemic, iSemantic 2021*, 2021. doi: 10.1109/iSemantic52711.2021.9573221.
- [30] R. Ranftl, K. Lasinger, D. Hafner, K. Schindler, and V. Koltun, “Towards Robust Monocular Depth Estimation: Mixing Datasets for Zero-Shot Cross-Dataset Transfer,” *IEEE Trans Pattern Anal Mach Intell*, vol. 44, no. 3, pp. 1623–1637, Mar. 2022, doi: 10.1109/TPAMI.2020.3019967.
- [31] T. Li, Z. Li, Y. Mu, and J. Su, “Pedestrian multi-object tracking based on YOLOv7 and BoT-SORT,” in *Third International Conference on Computer Vision and Pattern Analysis (ICCPA 2023)*, L. Shen and G. Zhong, Eds., SPIE, Aug. 2023, p. 68. doi: 10.1117/12.2684256.
- [32] S. Nindam, S.-H. Na, and H. J. Lee, “MultiFusedNet: A Multi-Feature Fused Network of Pretrained Vision Models via Keyframes for Student Behavior Classification,” *Applied Sciences*, vol. 14, no. 1, p. 230, Dec. 2023, doi: 10.3390/app14010230.