Enhancing Motoric Impulsivity Detection in Children through Deep Learning and Body Keypoint Recognition

Fahmy F. Dalimarta [a,b,*], Pulung N. Andono [a], Moch. A. Soeleman [a], Zainal A. Hasibuan [a]

[a] *Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang, Indonesia*
[b] *Faculty of Engineering and Informatics, Universitas Muhammadiyah Tegal, Tegal, Indonesia*
Corresponding author: [*]fahmy@mhs.dinus.ac.id

Abstract—**Quantifying motoric impulsivity in pediatric settings is crucial for safeguarding children and for devising effective intervention strategies. Existing quantitative techniques, such as accelerometry, have been utilized to assess it, but they often prove insufficient for accurately differentiating impulsive movements from regular ones. Conventional assessment methods are frequently used and rely on subjective assessments, which hinders the accurate characterization of impulsive behavior. To address this research gap, our study introduced an innovative objective approach using computer vision and deep learning techniques. We utilized MediaPipe to track precise body movement data from a child. The data were then analyzed using a Bidirectional Long Short-Term Memory (Bi-LSTM) network to process sequential information and recognize patterns indicative of impulsivity. Our approach successfully distinguished impulsive movements, marked by rapid changes in position and inconsistent movement velocities, from typical behavioral patterns with an accuracy rate of 98.21%. This research demonstrates the effectiveness of combining computer vision and deep learning to measure motoric impulsivity more precisely and impartially than prevailing qualitative techniques. Our model quantifies behaviors, enabling the development of improved safety protocols and targeted interventions in educational and recreational settings. This research has broader implications, suggesting a framework for future studies on pediatric motion analysis and behavioral assessment.**

Keywords—**Motoric impulsivity; quantitative; Bi-LSTM; keypoints; pose estimator.**

I. INTRODUCTION

Motoric impulsivity, characterized by spontaneous uncontrolled movements and an inability to suppress reflexes, is a prevalent symptom observed in several psychological disorders, including Attention-Deficit Hyperactivity Disorder (ADHD) [1], [2], bipolar disorder [3], [4], and depressive disorders [5], [6]. This impulsive behavior typically manifests as restlessness, constant motion, or interruption of others, which can present significant challenges in social settings, particularly in children [7]. Such behavior can result in safety hazards and difficulties in various environments such as schools, homes, and playgrounds. Children who are highly physically active, including those with motoric impulsivity disorders such as ADHD, are at a greater risk of sustaining injuries. The prevalence of injuries among children with ADHD ranges from 3 to 7% throughout their school years. Notably, children who have been diagnosed with this condition are more susceptible to accidents and nearly twice as likely to experience injuries as their peers who do not have ADHD [8].

Despite the significant impact of motoric impulsivity on various aspects of child development and well-being, accurate and objective assessment of this complex behavior remains a critical challenge due to the limitations of existing qualitative and quantitative methods. Assessment of motoric impulsivity is of paramount importance for several reasons. First, identifying and intervening early in children's impulsive behaviors can have a profound impact on their academic performance, social development, and overall well-being [9], [10]. Furthermore, a deeper understanding of the underlying mechanisms and elements contributing to motoric impulsivity can inform the development of more effective therapeutic approaches and targeted interventions for related disorders such as ADHD [11].

Various techniques have been employed to address this issue, which can be classified into two groups: quantitative and qualitative. Quantitative techniques have been found to be effective in objectively assessing and characterizing

motoric impulsivity. Among these techniques, motion tracking and analysis using specialized equipment such as cameras or inertial sensors is one approach. This method allows for precise measurement and quantification of various movement features, including velocity, acceleration, and directional changes.

Optical motion capture is a well-established method that involves attaching markers to a subject's body and employing specialized cameras to track the movement of these markers over a period of time. Through this process, a digital representation of the subject's movements can be obtained, facilitating the quantification of the movement patterns and dynamics. Studies have utilized this technique to analyze various aspects of limb movements, such as the range of motion, velocity, and coordination, in individuals with impulsivity-related disorders [12], [13].

Wearable inertial sensors, including accelerometers and gyroscopes, have been used to directly track and quantify body movements. These sensors can be incorporated into specialized attire or attached to specific body segments to provide precise measurements of linear and angular movements [14], [15]. By analyzing the collected data, researchers can extract various kinematic features such as acceleration profiles, jerk (rate of change of acceleration), and movement trajectories, which may be indicative of impulsive behavior.

Although quantitative techniques have demonstrated potential in assessing and quantifying motoric impulsivity, they often rely on specialized equipment or arrangements, which can prove costly and potentially intrusive for subjects, particularly for children. Furthermore, these methods may encounter challenges in capturing the subtleties and contexts of movements because they principally concentrate on raw kinematic data without accounting for environmental or situational factors.

On the other hand, the evaluation of motoric impulsivity has traditionally relied on qualitative methods, primarily involving subjective observations and clinical assessments. These techniques typically involve trained professionals such as psychologists or clinicians, who observe and score an individual's behavior using predefined scales or checklists.

A commonly used strategy is the implementation of standardized assessment tools such as the Conners' Rating Scales [16] or the ADHD Rating Scale [17]. These instruments offer a systematic means of evaluating various facets of impulsive conduct, including motor restlessness, squirming, and difficulty in remaining in one place. Medical practitioners and caregivers assess the frequency and intensity of these behaviors through observations, thereby facilitating the detection and quantification of motoric impulsivity.

Observational coding systems are another qualitative technique that entails trained observers systematically recording and categorizing specific instances of impulsive motor behaviors in structured or naturalistic settings. These coding systems typically consist of detailed operational definitions and guidelines for identifying and coding different types of impulsive movements, such as fidgeting, restless movements, or sudden changes in body position [18].

Qualitative methods have played a significant role in deepening our understanding of motoric impulsivity; however, they have some limitations. For instance, these assessments can be susceptible to observer bias, in addition to being dependent on the experience and training of the observer. Moreover, the process of capturing and quantifying intricate and rapidly changing movement patterns is inherently complicated and can lead to inconsistencies [19]. Qualitative techniques typically involve extended and resource-intensive observation periods, which may limit their scalability and practicality in real-world scenarios.

Among the two methods mentioned, neither the quantitative nor the qualitative approach presents an ideal solution for identifying motor impulsivity. In both cases, there is a trade-off. However, recent advancements in computer vision and deep-learning methodologies present promising opportunities to address this gap. These technologies are used to process and analyze intricate visual data, including human movements [20], [21]. By analyzing movement patterns and dynamics in video recordings, computer models can capture various movement attributes, such as the estimated position of keypoints of body parts, and calculate velocity, acceleration, and changes in direction. This information can then be used to classify movements into different levels of motoric impulsivity, potentially offering more consistent and objective assessments than those relying solely on human observation.

Although computer vision techniques have demonstrated potential for movement analysis [22], their application in quantifying and classifying motoric impulsivity remains relatively unexplored. Current automated movement analysis approaches often struggle with the intricate nature of movement patterns and unique characteristics of individual movements, making it challenging to develop consistent analysis systems. Although previous studies have applied automated techniques to movement analysis in specific domains, such as assessing neurodevelopmental disorders in infants [23], forensic analysis [24], and psychology-based eye movement studies [25], crucial aspects of movement data, including velocity and acceleration, have not been thoroughly investigated in the context of motoric impulsivity assessment.

To bridge this research gap, this study presents a video-based model for predicting motoric impulsivity that employs the MediaPipe pose estimator and Bi-Directional Long Short-Term Memory (Bi-LSTM) [26] architecture. The model aims to objectively analyze children's motoric impulsivity by extracting keypoints from real-time movements in videos, then calculating their movement metrics, thereby overcoming the limitations of current qualitative assessments, and utilizing computer vision technology to differentiate between impulsive and regular movements.

Bi-LSTM networks, which integrate two types of Recurrent Neural Networks (RNNs), have been effectively utilized in various applications, such as gesture recognition [27], video detection [28], and behavior detection [29], [30], [31], [32]. The capacity of these networks to capture temporal dependencies and relationships in sequential data makes them particularly suitable for precise prediction and classification [33]. Consequently, they have the potential to enhance the motoric impulsivity classification.

This study aimed to develop a comprehensive, standardized methodology for analyzing motoric impulsivity, involving the following:
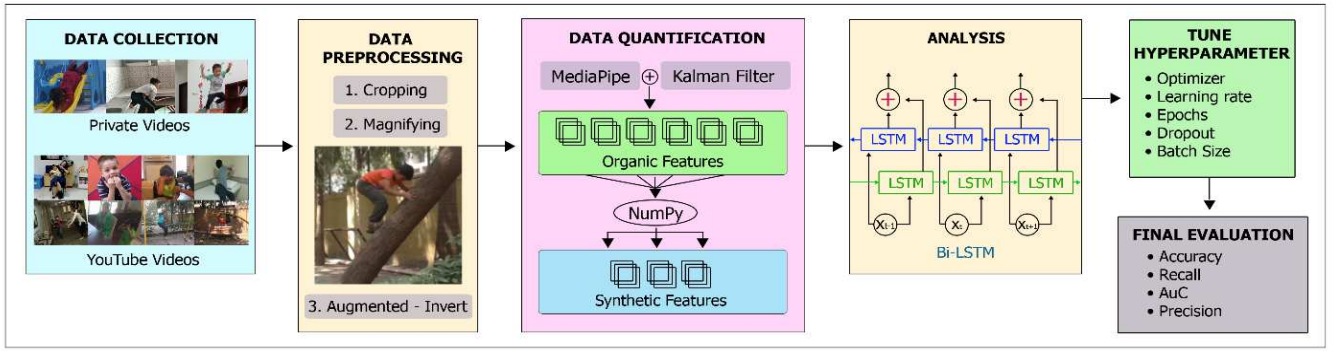
Fig. 1 The proposed methodology aims to classify motor-impulsive actions in videos.

a. Developing and evaluating a computer vision-based framework for accurately detecting and classifying motoric impulsivity levels from video data using the MediaPipe pose estimation model and Bi-LSTM architecture.

b. Investigating the impact of incorporating various movement features, such as velocity, acceleration, and directional changes, on the model's ability to distinguish impulsive movements from regular ones.

c. Assessing robustness and generalizability of the proposed approach by evaluating its performance for various movement quantification and subject characteristics.

d. Comparing the performance of the developed model to existing quantitative techniques, highlighting its potential advantages and limitations.

II. MATERIAL AND METHOD

Accurate assessment and quantification of impulsive movements pose significant challenges due to their complex and dynamic nature as well as the limitations of existing approaches. To address these challenges, we propose a novel framework that utilizes deep-learning techniques for precise movement analysis and classification. The following sections delve into the details of each stage, highlighting the innovative techniques and algorithms employed in our approach, which involves data collection, data preprocessing, feature extraction, model training, classification, and evaluation, as shown in Figure 1.

A. Data Collection

Two datasets are used in this study. The initial dataset comprised 66 videos obtained from YouTube, which showcased behaviors classified as motoric impulsiveness, such as repetitive, sudden, unnatural, and sporadic movements. These videos were obtained through the YouTube Data API by searching for ten pertinent keywords (motoric impulsiveness, disruptive behavior, snapping fingers, self-injury, scratching, tantrums, climbing, and twitching) and selecting the top six videos returned for each keyword. The dataset was expanded through a collection of twenty-five private videos from three children who had been diagnosed with ADHD-HI and displayed impulsive movements. The videos had to meet specific criteria, including a minimum duration of 10 seconds and the presence of observable motoric impulsive behavior. The data characteristics are presented in Table 1.

TABLE I
DATA COLLECTION CHARACTERISTICS

Type	Action	ID Videos	Duration	Snapshot
Private	Impulsive Motoric	01-PIM	21s	
Private	Impulsive Motoric	02-PIM	16s	
		10-PIM	10s	
		11-PIM	32s	
Private	Non-Impulsive	01-PNI	14s	
		02-PNI	27s	
		12-PNI	40s	
		13-PNI	22s	
YouTube	Impulsive Motoric	01-YIM	15s	
		02-YIM	21s	
		24-YIM	10s	
		25-YIM	13s	

Type	Action	ID Videos	Duration	Snapshot
YouTube	Non-Impulsive	01-YNI	28s	
		02-YNI	37s	
		40-YNI	19s	
		41-YNI	22s	

A total of 91 video clips were collected, and almost 40% showed individuals with psychological disorders exhibiting repetitive behaviors, unusual movements, or excessive activity. Each video was carefully annotated with metadata such as location context, perceived environmental stimuli, and duration of impulsivity. This thorough annotation process was conducted by the first author and a team of twelve kindergarten teachers and 6 elementary school teachers, aged between 20 and 39 years. Each video was annotated by the first author and at least two teachers. The process depicted in Figure 2 encompasses the entire video processing journey, culminating in the acquisition of numerical data that are subsequently fed into Bi-LSTM.
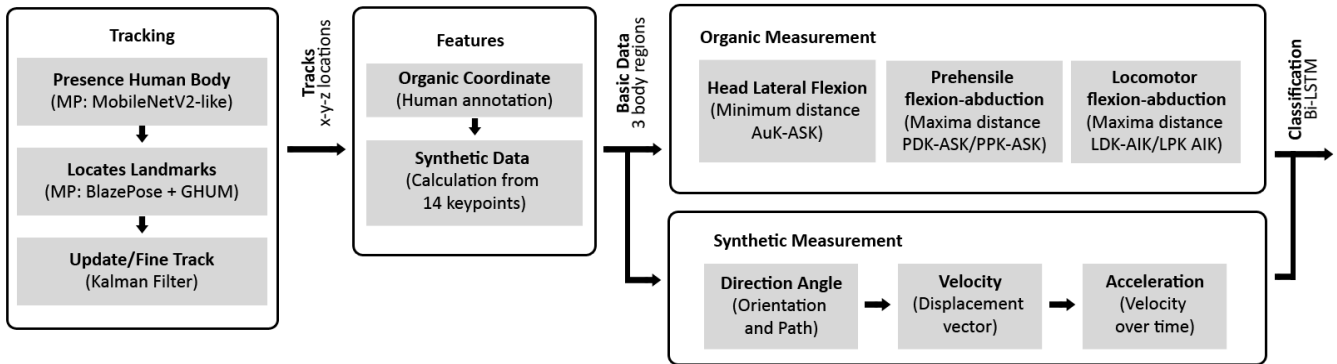

Fig. 2 Step-length methodology for data generation: Video processing produces x,y, and z-coordinates at each key point of the body. The coordinates are divided into organic and synthetic, and both are then processed respectively producing a complete dataset to be fed to Bi-LSTM.

B. Data Preprocessing

Of all 91 video sources, 37 videos displayed potential impulsive movements, and 54 videos displayed controlled movements. To prepare the dataset, all videos were cropped to leave the subject in the video, magnified to focus on the subject, and centered on the subject. All these processes were performed using Microsoft ClipChamp. Following the cropping process, the video has been transformed into a square format with a frame size of 540 x 540 pixels in dimensions. These steps result in a decrease in video quality and pixelation, as shown in Figure 3.

Fig. 3 Frames before preprocessing on the left and after preprocessing on the right.

To expand the dataset, a data augmentation process was implemented by reversing all videos while preserving their labels, ultimately doubling the dataset to 182 videos. The reversing process refers to flipping the image along one of its axes; this process was performed horizontally.

C. Data Quantification

MediaPipe uses the TensorFlow Lite library for building the model. It combines multiple models and offers a model-maker feature for transfer learning similar to TensorFlow. Our research used the MediaPipe Pose Landmarker to identify human body landmarks in the videos. The model processes a video feed to produce body-pose landmarks in three-dimensional space coordinates. Figure 4 illustrates the keypoint numbering on the human body.
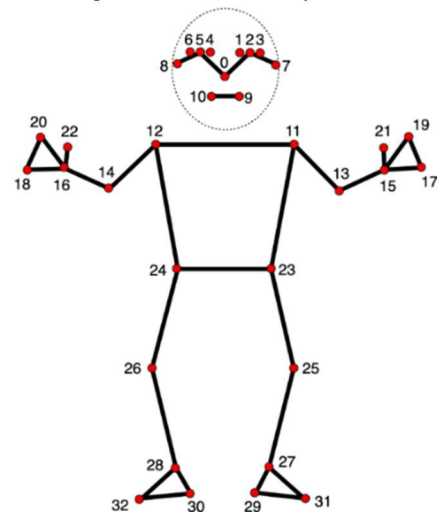

Fig. 4 The landmark pose model identified 33 crucial body keypoints, marking the approximate positions of various parts of the body [34]

278

This research examines the application of certain keypoints from MediaPipe to assess the movements. Specifically, we utilized only 14 out of the total 33 keypoints available in MediaPipe for this analysis. They are the left and right ears as auricle keypoints (AuK), the legs as locomotor keypoints (LK), and the hands as prehensile keypoints (PK). Several anchor keypoints (AK) were included as reference points for relative movement. The keypoints used in this study are listed in Table 2.

TABLE II
UTILIZED KEYPOINTS

Type	#KP	Name
Auricle (AuK)	7	L Ear
Auricle (AuK)	8	R Ear
Anchor Superior (ASK)	11	L Shoulder
Anchor Superior (ASK)	12	R Shoulder
Anchor Inferior (AIK)	23	L Collarbone
Anchor Inferior (AIK)	24	R Collarbone
Prehensile Proximal (PPK)	13	L Elbow
Prehensile Proximal (PPK)	14	R Elbow
Prehensile Distal (PDK)	15	L Wrist
Prehensile Distal (PDK)	16	R Wrist
Locomotor Proximal (LPK)	25	L Knee
Locomotor Proximal (LPK)	26	R Knee
Locomotor Distal (LDK)	27	L Ankle
Locomotor Distal (LDK)	28	R Ankle

During skeleton keypoint generation, MedaPipe frequently experiences intermittent location interference, which hinders their ability to identify objects accurately. We then used Kalman filters to overcome this limitation [35]. A total of 182

videos from the preprocessing stage were input into MediaPipe. Figure 5 shows the interim results of this process.

Fig. 5 Skeleton keypoints were overlaid on video clips processed by MediaPipe and then quantized.

The Kalman filter recursively estimates and predicts the system state based on noisy measurements, by initializing the initial keypoint position, predicting the next state for each frame, and correcting it using the observed keypoint. This process enables smooth keypoint tracking. For each video, we applied the following criteria to calculate the movement aspects, as illustrated in Figure 6.

a. Head lateral flexion: local minimum distance between the ear and shoulder keypoints (AuK-ASK).
b. Prehensile flexion-abduction: local maxima of the distance between the keypoints of the hand and shoulder (PDK-ASK/PPK-ASK).
c. Locomotor flexion-abduction: local maxima of the distance between the foot and hip keypoints (LDK-AIK/LPK-AIK).
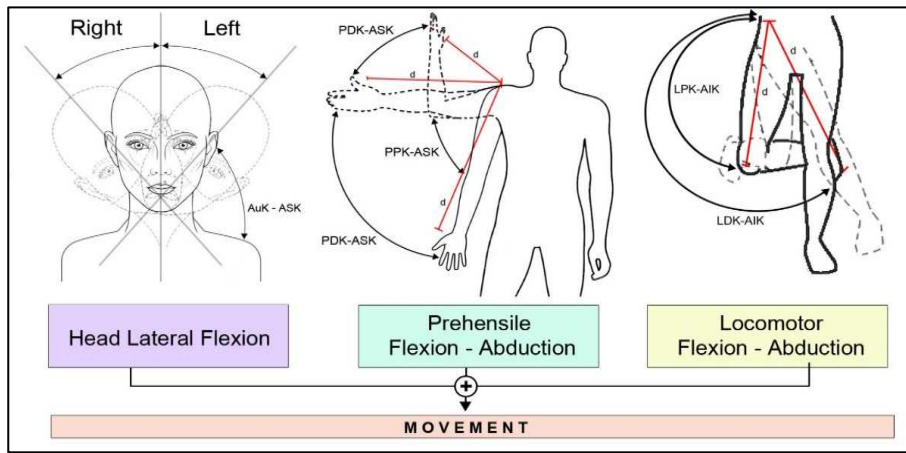d. The ground truth was manually set by frame-by-frame inspection.

Fig. 6 Three main criteria were used to calculate the movement aspects

MediaPipe recorded the keypoint data as comma-separated values (.csv) files. We separated the data into two types, based on the generation process. First, the organic features that contain the position of each keypoint per unit time are presented as three-dimensional [x,y,z] coordinates for the keypoints AuK, PPK, PDK, LPK, and LDK, calculated against the [0,0] frame point. These 3D data are automatically processed by MediaPipe's algorithms, which estimate 3D coordinates from 2D inputs using geometric principles, machine learning, and statistical methods. Second, synthetic keypoint data were derived by calculating the velocity, angle, acceleration, motion trajectory, head lateral flexion distance, prehensile flexion abduction, and locomotor flexion abduction from organic keypoint data.

D. Data Analysis and Evaluation

1) Movement Velocity Calculation: Movement velocity is a critical metric that can be used to differentiate between impulsive and controlled movement. In general, higher velocities tend to be indicative of impulsive behavior. Calculating the velocity of a keypoint's movement involves determining its initial and final positions, and the time it takes to move between them. Assuming 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 displacement 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, and $\Delta z = z_2 - z_1$ on the z axis. The displacement distance can be

calculated using the Pythagorean equation, which determines the displacement-vector length.

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

The following formula can be used to determine the velocity:

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

2) Movement Direction Calculation: The calculation of movement direction is of considerable importance, as it can offer valuable insights into the direction and orientation of movements, which may suggest specific impulsive behaviors or patterns. To determine the direction of the displacement, it is essential to compute the displacement vector and validate its direction. The displacement vector, which connects the initial position of the keypoint to its final position, is calculated using the following formula:

$$d = \Delta x \boldsymbol{i} + \Delta y \boldsymbol{j} + \Delta z \boldsymbol{k} \qquad (3)$$

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

$$\theta_x = \cos^{-1} \frac{\Delta x}{\sqrt{x^2 + y^2 + z^2}} \qquad (4)$$

$$\theta_y = \cos^{-1} \frac{\Delta y}{\sqrt{x^2 + y^2 + z^2}} \qquad (5)$$

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

3) Movement Acceleration Calculation: Determining the acceleration of a keypoint's movement in a three-dimensional space involves calculating the change in the vector velocity over a specific time frame. Acceleration monitoring can reveal sudden and jerky movements that are often indicative of impulsive behavior. By measuring the change in velocity per unit time, the initial and final velocities of the keypoint movement were determined prior to the calculation. Assuming that the initial velocity of the keypoint is $v_1 = v_{1x}\boldsymbol{i} + v_{1y}\boldsymbol{j} + v_{1z}\boldsymbol{k}$ and the final velocity is $v_2 = v_{2x}\boldsymbol{i} + v_{2y}\boldsymbol{j} + v_{2z}\boldsymbol{k}$ with $v_{1x}, v_{1y}, v_{1z}, v_{2x}, v_{2y}, v_{2z}$ are the components of the velocity on 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 $\Delta v_x = v_{2x} - v_{1x}$ whereas on the y-axis the formula applied $\Delta v_y = v_{2y} - v_{1y}$ and on the z-axis $\Delta v_z = v_{2z} - v_{1z}$. The acceleration of motion is the change in the total velocity divided by the total time. This was calculated using the following equation:

$$a = \frac{\Delta v_x}{\Delta t}\boldsymbol{i} + \frac{\Delta v_y}{\Delta t}\boldsymbol{j} + \frac{\Delta v_z}{\Delta t}\boldsymbol{k} \qquad (7)$$

4) Data Analysis: Our model was implemented by allocating the processed dataset into three segments: 60% for training, 20% for validation, and 20% for testing. Table 3 shows 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. A confusion matrix is used to evaluate the performance of the algorithm by comparing the predicted and actual class instances. True Positive, True Negative, False Positive, and False Negative values were determined. Our research employed metrics, such as Accuracy, Sensitivity (recall), Precision, and AUC, to assess the algorithm. This metric was calculated as follows:

$$Accuration = \frac{(TP + TN)}{(TP + FP + TN + FN)} \qquad (8)$$

$$Precision = \frac{(TP)}{(TP + FP)} \qquad (9)$$

$$Recall = \frac{(TP)}{(TP + FN)} \qquad (10)$$

TABLE III
BI-LSTM HYPERPARAMETER CONFIGURATION

Hyperparameter	Value
Activation function	Sigmoid
Loss function	Binary Cross entropy
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

III. RESULTS AND DISCUSSION

Our innovative computer vision-based framework demonstrates significant advancements over the existing methods for quantifying and classifying motoric impulsivity. We assessed the effectiveness of our approach using a comprehensive dataset comprising videos from various sources, including online platforms and specially recorded sessions. The dataset was carefully curated to capture a wide range of impulsive movements exhibited by children in various settings, ensuring a thorough evaluation of our model's performance.

A. Movement Detection and Quantification Results

1) Head Lateral Flexion: Head lateral flexion, also known as head side-to-side movement, involves bending the head towards the left or right shoulder. This movement is often associated with impulsive behaviors such as fidgeting or restlessness and can be a valuable indicator of motoric impulsivity. By assessing head lateral flexion using frame-by-frame video analysis and tracking specific ear landmarks, we can quantify the degree and frequency of these movements. This metric is practically significant, as excessive or repetitive head lateral flexion can disrupt focus, interfere with activities, and potentially indicate underlying conditions, such as ADHD or anxiety disorders. Table 4 lists the quantified results for the head lateral flexion sections of twenty-five private videos.

2) Prehensile flexion-abduction: Prehensile flexion-abduction refers to a complex wrist movement often used in

grasping objects. Tracking this metric is important because abnormal or excessive wrist movements can be indicative of impulsive behaviors, such as fidgeting, repetitive motions, or self-stimulatory behaviors. By carefully monitoring the wrist joint angle and establishing a neutral zero position, we can quantify the degree and range of prehensile flexion-abduction movements. This information is relevant for identifying potential impulsive behaviors related to hand and arm movements that can interfere with daily activities, social interactions, or learning environments. Table 5 shows the results of quantifying movements in the prehensile flexion-abduction section.

3) Locomotor *flexion-abduction:* Locomotor flexion-abduction refers to complex flexion and abduction movements of the lower extremity joints during activities such as walking or running. Quantifying these movements is crucial because they can reveal gait patterns, coordination difficulties, or excessive movements associated with impulsivity. By tracking anatomical landmarks and calculating joint angles over time, the overall range of motion and coordination of leg and foot movements can be assessed. Table 6 shows the results of quantifying movements in the locomotor flexion-abduction section.

TABLE IV
THE MOVEMENT QUANTIFICATION OF HEAD LATERAL FLEXION SYNTHETIC DATA

Video	$\bar{d}$ AuK-ASK		$\bar{\theta}$ AuK		$\bar{v}$ AuK		$\bar{a}$ AuK	
	L	R	L	R	L	R	L	R
# 1	0.0034	0.0083	26.48	55.29	0.016	0.108	0.021	0.089
# 2	0.0028	0.0067	24.95	48.85	0.022	0.083	0.023	0.070
# 3	0.0035	0.0070	40.45	65.38	0.014	0.096	0.012	0.070
# 4	0.0028	0.0116	22.15	68.72	0.019	0.238	0.021	0.190
# 5	0.0046	0.0089	8.36	32.87	0.038	0.086	0.112	0.129
# 6	0.0064	0.0113	14.53	44.79	0.038	0.097	0.049	0.098
# 7	0.0096	0.0164	18.86	33.16	0.056	0.111	0.090	0.134
# 8	0.0018	0.0055	38.59	59.94	0.011	0.075	0.011	0.063
# 9	0.0024	0.0077	39.70	61.04	0.016	0.097	0.017	0.073
# 10	0.0022	0.0071	71.27	67.21	0.027	0.121	0.016	0.086
# 11	0.0031	0.0194	21.03	53.53	0.028	0.325	0.019	0.264
# 12	0.0035	0.0095	33.18	58.86	0.019	0.139	0.029	0.112
# 13	0.0050	0.0105	26.08	49.19	0.026	0.139	0.038	0.113
# 14	0.0018	0.0035	29.76	60.73	0.022	0.067	0.016	0.051
# 15	0.0022	0.0060	11.61	50.40	0.019	0.064	0.016	0.055
# 16	0.0060	0.0113	18.69	46.79	0.028	0.132	0.043	0.113
# 17	0.0012	0.0030	62.68	54.43	0.009	0.024	0.011	0.026
# 18	0.0041	0.0102	19.09	51.13	0.031	0.127	0.047	0.112
# 19	0.0058	0.0085	33.07	50.87	0.024	0.076	0.022	0.065
# 20	0.0053	0.0112	36.60	47.34	0.043	0.164	0.035	0.119
# 21	0.0030	0.0091	28.15	58.21	0.026	0.127	0.031	0.107
# 22	0.0058	0.0246	26.49	61.53	0.039	0.382	0.039	0.290
# 23	0.0038	0.0115	24.29	57.21	0.025	0.173	0.031	0.148
# 24	0.0036	0.0070	38.73	56.38	0.014	0.058	0.032	0.060
# 25	0.0021	0.0054	58.44	68.38	0.012	0.078	0.012	0.060

TABLE V
THE MOVEMENT QUANTIFICATION OF PREHENSILE FLEXION-ABDUCTION SYNTHETIC DATA

Video	$\bar{d}$ PPK-ASK		$\bar{\theta}$ PPK		$\bar{v}$ PPK		$\bar{a}$ PPK		$\bar{d}$ PDK-ASK		$\bar{\theta}$ PDK		$\bar{v}$ PDK		$\bar{a}$ PDK	
	L	R	L	R	L	R	L	R	L	R	L	R	L	R	L	R
# 1	0.0051	0.0115	49.80	55.37	0.043	0.115	0.042	0.096	0.5011	0.0230	1.01	59.39	0.059	0.168	0.055	0.138
# 2	0.0042	0.0092	38.02	54.50	0.038	0.090	0.035	0.072	0.3761	0.0177	1.20	58.19	0.056	0.125	0.045	0.098
# 3	0.0026	0.0126	62.98	67.36	0.025	0.115	0.017	0.085	0.6295	0.0280	0.33	69.26	0.018	0.212	0.012	0.152
# 4	0.0052	0.0110	34.07	69.79	0.034	0.196	0.036	0.154	0.6749	0.0209	0.62	67.68	0.050	0.229	0.039	0.180
# 5	0.0083	0.0146	16.16	41.72	0.050	0.083	0.086	0.089	0.5139	0.0277	0.79	53.69	0.065	0.152	0.067	0.128
# 6	0.0081	0.0165	29.87	47.99	0.065	0.121	0.065	0.107	0.5695	0.0339	1.57	45.65	0.100	0.199	0.080	0.179
# 7	0.0117	0.0169	15.43	38.65	0.070	0.123	0.066	0.119	0.3905	0.0281	2.00	42.34	0.098	0.142	0.073	0.142
# 8	0.0023	0.0079	51.73	62.04	0.021	0.073	0.015	0.064	0.6145	0.0172	0.42	60.41	0.030	0.111	0.023	0.094
# 9	0.0030	0.0098	51.26	64.15	0.024	0.097	0.018	0.076	0.5580	0.0199	0.81	63.65	0.044	0.127	0.036	0.110
# 10	0.0024	0.0112	68.48	73.87	0.032	0.135	0.021	0.095	0.3560	0.0200	0.52	72.53	0.022	0.183	0.017	0.131
# 11	0.0047	0.0318	25.23	49.53	0.043	0.197	0.030	0.180	0.5340	0.0533	0.80	53.52	0.054	0.340	0.034	0.286
# 12	0.0055	0.0155	53.64	64.22	0.034	0.149	0.035	0.120	0.4842	0.0316	0.65	67.75	0.034	0.238	0.035	0.189
# 13	0.0065	0.0149	40.12	49.48	0.040	0.114	0.042	0.103	0.4939	0.0316	0.97	54.63	0.051	0.205	0.049	0.176
# 14	0.0030	0.0069	47.52	68.26	0.025	0.072	0.017	0.054	0.4917	0.0147	0.46	74.33	0.029	0.117	0.018	0.086
# 15	0.0023	0.0062	30.19	46.92	0.022	0.056	0.012	0.049	0.4609	0.0138	0.49	54.57	0.027	0.094	0.015	0.081
# 16	0.0066	0.0141	37.92	50.27	0.063	0.131	0.064	0.116	0.5339	0.0317	1.58	58.49	0.079	0.215	0.076	0.177
# 17	0.0022	0.0055	73.03	54.60	0.012	0.035	0.013	0.036	0.3905	0.0107	0.32	54.62	0.016	0.061	0.014	0.059
# 18	0.0066	0.0179	35.28	61.42	0.042	0.154	0.048	0.131	0.7107	0.0359	0.87	64.50	0.074	0.235	0.069	0.194
# 19	0.0109	0.0195	50.11	56.99	0.093	0.165	0.070	0.136	0.7985	0.0428	0.98	64.64	0.082	0.301	0.066	0.238
# 20	0.0066	0.0252	56.66	56.32	0.067	0.180	0.045	0.140	0.5183	0.0590	1.48	63.69	0.085	0.386	0.063	0.285
# 21	0.0045	0.0143	43.66	56.11	0.034	0.109	0.033	0.093	0.5168	0.0286	0.91	63.25	0.052	0.180	0.043	0.146
# 22	0.0095	0.0219	36.76	64.21	0.061	0.269	0.055	0.203	0.3684	0.0384	2.31	70.01	0.092	0.337	0.083	0.246
# 23	0.0053	0.0121	43.07	54.78	0.041	0.136	0.044	0.121	0.4741	0.0214	1.11	56.17	0.063	0.170	0.058	0.145
# 24	0.0059	0.0145	43.22	62.93	0.036	0.112	0.049	0.100	0.4053	0.0305	4.32	63.68	0.055	0.185	0.056	0.159
# 25	0.0038	0.0105	57.12	66.64	0.026	0.081	0.023	0.065	0.4606	0.0231	0.81	71.28	0.047	0.151	0.043	0.116

| Video	$\bar{d}$ LPK-AIK		$\bar{\theta}$ LPK		$\bar{v}$ LPK		$\bar{a}$ LPK		$\bar{d}$ LDK-AIK		$\bar{\theta}$ LDK		$\bar{v}$ LDK		$\bar{a}$ LDK	
	L	R	L	R	L	R	L	R	L	R	L	R	L	R	L	R
# 1	0.0081	0.0233	43.72	55.71	0.049	0.100	0.042	0.086	0.0303	0.0354	59.42	58.37	0.177	0.170	0.144	0.146
# 2	0.0073	0.0204	35.64	63.30	0.034	0.096	0.032	0.072	0.0258	0.0307	65.20	69.65	0.151	0.166	0.114	0.121
# 3	0.0035	0.0233	56.19	65.98	0.024	0.103	0.015	0.076	0.0367	0.0412	69.24	69.02	0.249	0.229	0.176	0.163
# 4	0.0077	0.0393	36.34	45.06	0.035	0.084	0.034	0.076	0.0434	0.0574	64.33	63.54	0.242	0.198	0.199	0.162
# 5	0.0122	0.0266	20.06	42.78	0.051	0.079	0.063	0.083	0.0533	0.0443	54.60	53.94	0.226	0.171	0.196	0.146
# 6	0.0083	0.0338	35.52	44.99	0.050	0.119	0.052	0.121	0.0439	0.0539	56.83	53.33	0.235	0.231	0.205	0.214
# 7	0.0152	0.0373	15.43	47.08	0.077	0.163	0.083	0.133	0.0448	0.0542	51.22	52.06	0.231	0.244	0.163	0.195
# 8	0.0071	0.0407	58.40	71.32	0.043	0.238	0.034	0.180	0.0604	0.0651	77.26	73.48	0.414	0.400	0.300	0.295
# 9	0.0039	0.0244	55.89	63.41	0.027	0.112	0.018	0.089	0.0358	0.0361	72.94	63.05	0.228	0.176	0.171	0.133
# 10	0.0069	0.0283	76.58	67.51	0.038	0.147	0.024	0.100	0.0339	0.0374	74.56	74.63	0.250	0.202	0.169	0.138
# 11	0.0049	0.0728	33.64	54.38	0.039	0.374	0.023	0.281	0.1090	0.1226	53.70	59.13	0.654	0.673	0.500	0.499
# 12	0.0088	0.0369	43.91	66.82	0.037	0.177	0.036	0.138	0.0470	0.0579	70.21	70.07	0.288	0.310	0.225	0.235
# 13	0.0101	0.0428	46.69	63.10	0.049	0.221	0.044	0.175	0.0624	0.0664	67.25	66.44	0.411	0.387	0.304	0.283
# 14	0.0035	0.0120	46.56	65.42	0.022	0.050	0.011	0.038	0.0165	0.0186	58.14	70.41	0.097	0.087	0.071	0.065
# 15	0.0034	0.0236	43.55	60.82	0.025	0.150	0.015	0.108	0.0360	0.0388	66.12	68.00	0.245	0.240	0.171	0.174
# 16	0.0102	0.0328	32.48	51.85	0.056	0.158	0.059	0.134	0.0493	0.0538	59.86	58.90	0.292	0.290	0.239	0.235
# 17	0.0027	0.0072	78.44	56.82	0.011	0.030	0.012	0.028	0.0121	0.0141	65.33	64.26	0.066	0.073	0.057	0.060
# 18	0.0106	0.0337	37.69	60.38	0.049	0.165	0.050	0.136	0.0487	0.0553	63.59	66.07	0.290	0.302	0.230	0.233
# 19	0.0071	0.0266	58.30	64.77	0.033	0.123	0.020	0.096	0.0500	0.0464	69.38	65.12	0.311	0.250	0.233	0.191
# 20	0.0110	0.0391	37.10	56.16	0.060	0.146	0.046	0.119	0.0465	0.0603	56.60	59.95	0.262	0.275	0.202	0.217
# 21	0.0071	0.0386	35.11	65.67	0.038	0.210	0.039	0.159	0.0592	0.0634	71.95	72.71	0.386	0.393	0.282	0.275
# 22	0.0139	0.0598	44.32	62.12	0.066	0.172	0.054	0.126	0.0507	0.0894	68.18	75.57	0.305	0.362	0.234	0.260
# 23	0.0090	0.0348	45.55	55.89	0.044	0.132	0.044	0.112	0.0448	0.0568	66.96	65.12	0.275	0.265	0.214	0.208
# 24	0.0055	0.0208	52.11	59.47	0.025	0.078	0.034	0.072	0.0228	0.0288	62.00	62.33	0.121	0.125	0.105	0.109
# 25	0.0062	0.0258	56.24	68.40	0.035	0.122	0.030	0.098	0.0372	0.0450	76.14	76.49	0.247	0.253	0.182	0.190
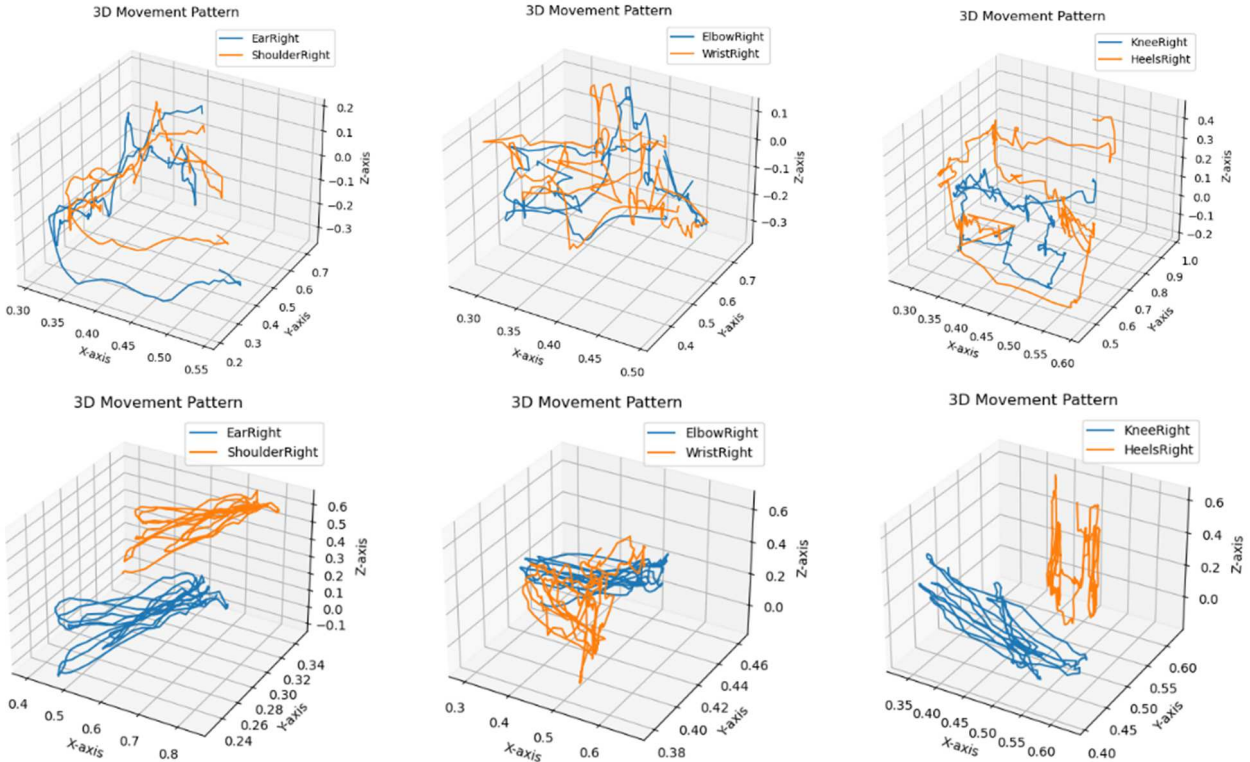
Fig. 7 The 3D environment displays the movement patterns of AuK, PPK, and LPK. On the upside are impulsive-motoric subjects, while the downside displays non-impulsive subjects.

We transformed the raw data into a 3D visual representation starting with 25 videos, each containing distinct numerical features. Figure 7 shows the two samples selected for visualization. By placing these contrasting datasets side by side on a 3D landscape, we gain a comprehensive view of their dual nature, highlighting the differences between motor impulsiveness and non-impulsiveness.

B. Overall Performance

We conducted a thorough evaluation of our proposed model using each of the quantification sub-datasets presented

above. The optimal configuration of the model is listed in Table 3. Our model exhibited impressive performance on the test dataset, achieving an accuracy of 98.21%. This metric reflects the general effectiveness of the model in classifying the sample and assesses the proportion of correct predictions relative to the total number of observations, as shown in Figure 8.

Figure 9. displays the AUC value of our classification model, which was 0.978. This result indicates that our model performed significantly better than random chance, as indicated by the dashed diagonal line on the curve. In addition, our model demonstrated the ability to classify instances correctly. The AUC value of 0.978 was excellent, indicating a high level of accuracy and reliability in classification predictions.

The performance of our classification model is depicted by the precision-recall curve in Figure 10, illustrating its capabilities at various decision thresholds. Precision refers to the ratio of true positives to the sum of true positives and false positives, indicating the accuracy of positive predictions. Recall represents the ratio of true positives to the sum of true positives and false negatives, emphasizing the model's ability to identify all positive events.

The precision-recall curve shows the interaction between precision and recall when the decision threshold is varied. Models with perfect precision and recall are marked in the upper-right corner. Our curve demonstrated excellent results with satisfactory recall when precision was high. The area under the curve (AUC-PR) of 0.99 indicates a good balance between precision and recall, implying that the model effectively makes positive predictions while capturing actual positive events.
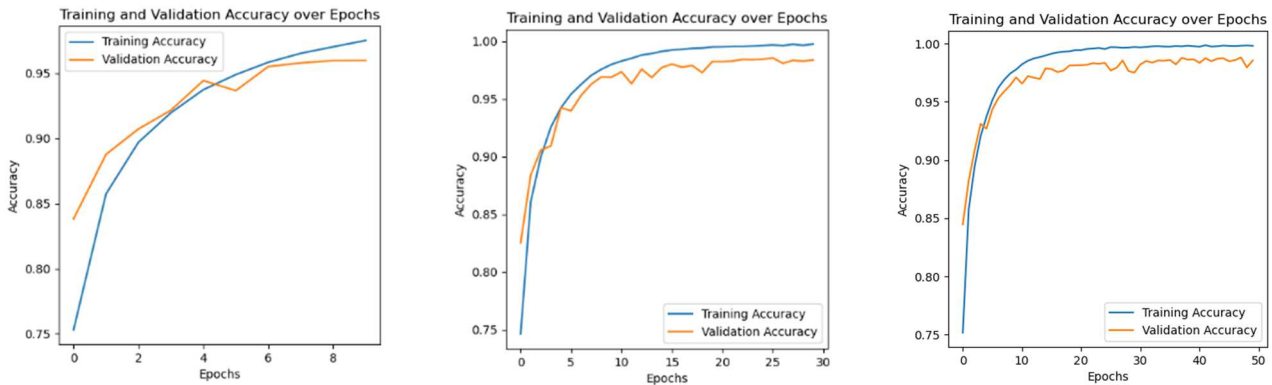

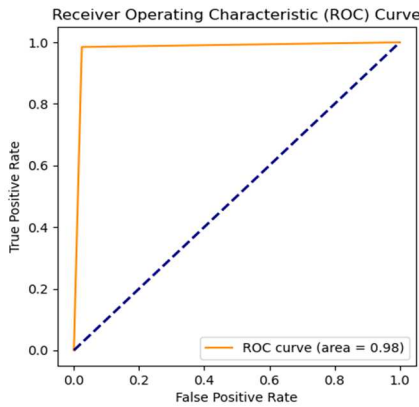Fig. 8 Accuracy curves between training and validation for 10, 30, and 50-epochs

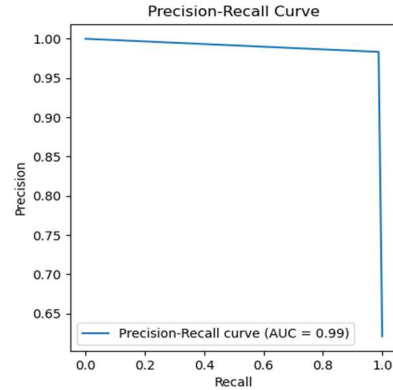Fig. 9 RoC curves using threshold 0.5 on the datasets

Fig. 10 Precision-recall curves for training and testing for 50-epochs.

TABLE VII
ACCURACY COMPARISON WITH PREVIOUS STUDY

Reference	Method Used	Description of the Experiment	Accuracy
[30]	R-CNN, MediaPipe, LSTM	Using deep learning techniques to detect and classify the behavior of manufacturing workers based on their poses and interactions with objects in the environment. This was done to identify normal and anomalous behavior, with the goal of improving safety and productivity in manufacturing settings.	94%
[36]	CNN, SP-CNN	The experiments aimed to evaluate the performance of the proposed SP-CNN approach in real-time performance, security, and abnormal behavior detection using a Jetson TX2 device and a cloud server. The performance of SP-CNN was evaluated based on its real-time performance, security, and detection accuracy using a K-nearest neighbor (KNN) classifier.	Walk behavior: 96.46% Fall behavior: 98.61%

Reference	Method Used	Description of the Experiment	Accuracy
Our Research	MediaPipe Bi-LSTM	Assessing motor impulsivity in children's computer vision and deep learning. The model, incorporating MediaPipe and a Bi-LSTM architecture. This highlights the effectiveness of combining these technologies for precise motor impulsivity analysis, valuable for enhancing safety measures.	98, 21%

The confusion matrix revealed that the model rarely mislabeled non-impulsive movements as impulsive (recall 99.07%) while capturing impulsive movements (sensitivity 97.46%). This demonstrates its ability to differentiate between true impulsive movements and regular movements. Collectively, the performance metrics, including accuracy, precision, recall, and F1 scores, indicate a robust, high-performance model for accurately categorizing instances across different classes. In the proposed work, significant progress was achieved compared with similar architectures, as shown in Table 7. This table highlights the differences between the proposed design and those presented in [30] and [36].

However, our study had certain limitations that need to be addressed. Accurately interpreting motoric impulsivity requires consideration of the environmental context in which the activity or movement occurs. Relying solely on predefined definitions may not capture the appropriateness of observed movements within a given context. In addition, the current implementation of MediaPipe is limited to singular subjects and cannot be simultaneously applied to multiple subjects.

To enhance the generalizability and robustness of our approach, future research should focus on integrating environmental context information into the analysis pipeline, potentially through scene understanding or situational awareness modeling. This would provide a more comprehensive understanding of the contextual factors influencing the observed behaviors. Furthermore, extending the pose estimation and analysis capabilities to handle multiple subjects simultaneously would broaden the applicability of our method to various real-world scenarios, such as classrooms, playgrounds, or group therapy sessions.

Additionally, exploring the integration of multimodal data sources, such as physiological signals or environmental factors, could provide a more holistic understanding of motoric impulsivity and its triggers, leading to personalized and effective interventions.

IV. CONCLUSION

Our study demonstrates the efficacy of computer vision and deep learning techniques in accurately identifying and quantifying motoric impulsivity in children. By employing MediaPipe pose estimation and the Bi-LSTM architecture, we achieved an accuracy rate of 98.21% in distinguishing impulsive movements from regular ones, leveraging features such as abrupt changes in body position, erratic velocity/acceleration, and recurring motions. These findings highlight the potential of automated systems to provide objective and precise assessments of motoric impulsivity, thereby overcoming the limitations of the subjective methodologies.

The real-world application of the proposed approach has practical implications, such as designing classrooms that minimize distractions, promoting focus, and implementing safety measures in playgrounds to reduce accidents caused by motoric impulsivity. Our method can also help to create supportive home environments by identifying triggers for impulsive behaviors and suggesting modifications, thus enabling caregivers to foster a nurturing environment that supports their children's well-being and development.

Our work not only focuses on environmental design but also paves the way for a paradigm shift in pediatric psychology and developmental neuroscience, enabling data-driven precision and personalized interventions for childhood impulsivity. While our study yielded promising results, it is crucial to acknowledge its limitations, such as the limited sample size and focus on preteens. Future research should explore larger and more diverse datasets and investigate the generalizability of this approach to different age groups and disorders.

NOMENCLATURE

$\bar{d}$	mean of distance	kpx
$\bar{\theta}$	mean of angle	rad
$\bar{v}$	mean of velocity	kpx/s
$\bar{a}$	mean of acceleration	kpx/s^2

ACKNOWLEDGMENT

REFERENCES

[1] J. C. Corona, "Role of Oxidative Stress and Neuroinflammation in Attention-Deficit/Hyperactivity Disorder," *Antioxidants*, vol. 9, no. 11, p. 1039, Oct. 2020, doi: 10.3390/antiox9111039.

[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/fpsyt.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] T. M. K. ElSehrawy, E. A. Elela, G. A. M. Hassan, M. El Missiry, S. A. Nabi, and M. F. Soliman, "A study of emotional intelligence in an Egyptian sample of offspring of patients with schizophrenia," *Middle East Current Psychiatry*, vol. 29, no. 1, p. 48, Dec. 2022, doi:10.1186/s43045-022-00216-x.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] J. Pereira, P. Vagos, A. Fonseca, H. Moreira, M. C. Canavarro, and D. Rijo, "The Children's Revised Impact of Event Scale: Dimensionality and Measurement Invariance in a Sample of Children and Adolescents Exposed to Wildfires," *J Trauma Stress*, vol. 34, no. 1, pp. 35–45, Feb. 2021, doi: 10.1002/jts.22634.

[10] B. Wolff, E. Sciberras, J. He, G. Youssef, V. Anderson, and T. J. Silk, "The Role of Sleep in the Relationship Between ADHD Symptoms and Stop Signal Task Performance," *J Atten Disord*, vol. 25, no. 13, pp. 1881–1894, Nov. 2021, doi: 10.1177/1087054720943290.

[11] J. C. Vázquez, O. Martin de la Torre, J. López Palomé, and D. Redolar-Ripoll, "Effects of Caffeine Consumption on Attention Deficit Hyperactivity Disorder (ADHD) Treatment: A Systematic Review of Animal Studies," *Nutrients*, vol. 14, no. 4, p. 739, Feb. 2022, doi:10.3390/nu14040739.

[12] G. Abbadessa *et al.*, "Digital therapeutics in neurology," *J Neurol*, vol. 269, no. 3, pp. 1209–1224, Mar. 2022, doi: 10.1007/s00415-021-10608-4.

[13] I. Weygers, M. Kok, M. Konings, H. Hallez, H. De Vroey, and K. Claeys, "Inertial Sensor-Based Lower Limb Joint Kinematics: A Methodological Systematic Review," *Sensors*, vol. 20, no. 3, p. 673, Jan. 2020, doi: 10.3390/s20030673.

[14] 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.

[15] 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.

[16] 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.

[17] G. J. DuPaul, T. J. Power, A. D. Anastopoulos, and R. C. Reid, "Adhd Rating Scale-IV: Checklists, Norms, and Clinical Interpretation," 1998.

[18] G. J. DuPaul and G. Stoner, *ADHD in the schools: Assessment and intervention strategies, 3rd ed.* New York, NY, US: The Guilford Press, 2014.

[19] J. Barth, J. W. Klaesner, and C. E. Lang, "Relationships between accelerometry and general compensatory movements of the upper limb after stroke," *J Neuroeng Rehabil*, vol. 17, no. 1, p. 138, Dec. 2020, doi: 10.1186/s12984-020-00773-4.

[20] M. S. Islam *et al.*, "Using AI to measure Parkinson's disease severity at home," *NPJ Digit Med*, vol. 6, no. 1, p. 156, Aug. 2023, doi:10.1038/s41746-023-00905-9.

[21] K. Luxem *et al.*, "Open-source tools for behavioral video analysis: Setup, methods, and best practices," *Elife*, vol. 12, Mar. 2023, doi:10.7554/eLife.79305.

[22] H. Imai *et al.*, "A lack of specific motor patterns between rhythmic/non-rhythmic masticatory muscle activity and bodily movements in sleep bruxism," *J Prosthodont Res*, vol. 65, no. 3, p. JPR_D_20_00012, 2021, doi: 10.2186/jpr.JPR_D_20_00012.

[23] W. Baccinelli *et al.*, "Movidea: A Software Package for Automatic Video Analysis of Movements in Infants at Risk for Neurodevelopmental Disorders," *Brain Sci*, vol. 10, no. 4, p. 203, Mar. 2020, doi: 10.3390/brainsci10040203.

[24] A. G. Boyarov, O. O. Vlasov, and I. S. Siparov, "Methodology for Determining Time Intervals by Video Recordings," *Theory and Practice of Forensic Science*, vol. 17, no. 2, pp. 58–69, Aug. 2022, doi: 10.30764/1819-2785-2022-2-58-69.

[25] J. H. Hsiao, H. Lan, Y. Zheng, and A. B. Chan, "Eye movement analysis with hidden Markov models (EMHMM) with co-clustering," *Behav Res Methods*, vol. 53, no. 6, pp. 2473–2486, Dec. 2021, doi:10.3758/s13428-021-01541-5.

[26] 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.

[27] Ü. Atila and F. Sabaz, "Turkish lip-reading using Bi-LSTM and deep learning models," *Engineering Science and Technology, an International Journal*, vol. 35, p. 101206, Nov. 2022, doi:10.1016/j.jestch.2022.101206.

[28] 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.

[29] 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.

[30] R. Rijayanti, M. Hwang, and K. Jin, "Detection of Anomalous Behavior of Manufacturing Workers Using Deep Learning-Based Recognition of Human–Object Interaction," *Applied Sciences*, vol. 13, no. 15, p. 8584, Jul. 2023, doi: 10.3390/app13158584.

[31] 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/ijsse.120510.

[32] 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.

[33] G. Van Houdt, C. Mosquera, and G. Nápoles, "A review on the long short-term memory model," *Artif Intell Rev*, vol. 53, no. 8, pp. 5929–5955, Dec. 2020, doi: 10.1007/s10462-020-09838-1.

[34] Google, "MediaPipe Pose Landmark." Accessed: Jan. 05, 2024. [Online]. Available: https://developers.google.com/mediapipe/solutions/vision/pose_land marker

[35] 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 Creatives for Digital Innovation and Communication within Global Pandemic, iSemantic 2021*, 2021. doi:10.1109/iSemantic52711.2021.9573221.

[36] J. Qiu, X. Yan, W. Wang, W. Wei, and K. Fang, "Skeleton-Based Abnormal Behavior Detection Using Secure Partitioned Convolutional Neural Network Model," *IEEE J Biomed Health Inform*, vol. 26, no. 12, pp. 5829–5840, Dec. 2022, doi:10.1109/JBHI.2021.3137334.