IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Mathematical Modelling of Engineering Problems
Vol. 12, No. 10, October, 2025, pp. 3718-3728
Journal homepage: http://iieta.org/journals/mmep

Impact of Image Enhancement on Deep Learning-Based Recognition of Activity Prompts in Children with Autism Using Motion History Images

Indah Werdiningsih[1,2], Ira Puspitasari[2,3*], Rimuljo Hendradi[2]

[1] Doctoral Program of Mathematics and Natural Sciences, Faculty of Science and Technology, Universitas Airlangga, Surabaya 60115, Indonesia
[2] Information Systems Study Program, Faculty of Science and Technology, Universitas Airlangga, Surabaya 60115, Indonesia
[3] Research Center for Quantum Engineering Design, Faculty of Science and Technology, Universitas Airlangga, Surabaya 60115, Indonesia

Corresponding Author Email: ira-p@fst.unair.ac.id

https://doi.org/10.18280/mmep.121035

Received: 28 August 2025
Revised: 8 October 2025
Accepted: 17 October 2025
Available online: 31 October 2025

Keywords:
assistive technology, ASD, deep learning, image enhancement, MHI, prompt recognition

ABSTRACT

Children with Autism Spectrum Disorder (ASD) often struggle with activity recognition and routine following, requiring continuous assistance that is not always available. Automated recognition of activity prompts could provide support for independent daily functioning and enable personalized therapeutic interventions. This work focuses on enhancing robustness under complex visual conditions while maintaining real-time computational feasibility. The proposed approach introduces a novel integration of Contrast Limited Adaptive Histogram Equalization (CLAHE) for contrast optimization and Motion History Image (MHI) for temporal representation within Convolutional Neural Network (CNN) architectures, namely VGG19 and MobileNetV2. The dataset consists of 1.083 videos of eating and drinking activities, categorized by physical, gesture, and verbal prompts. CLAHE improves visual clarity, yielding an Means Square Error (MSE) of 10.586, a Peak Signal to Noise Ratio (PSNR) of 38.275, and an Structural Similarity Index Measure (SSIM) of 0.997, indicating enhanced image quality. The proposed model achieved an accuracy of 73.96% with a computation time of 308 seconds, compared to 72.92% accuracy and 420 seconds without enhancement. While the integration of CLAHE and MHI with VGG19 enhances computational efficiency, accuracy improvements are modest due to the dataset's inherent complexity. These findings highlight that integrating motion-based features with image enhancement supports practical real-time deployment of assistive technologies.

1. INTRODUCTION

Autism Spectrum Disorder (ASD) is a mental and neurological disease in children that affects social interaction, communication, and behavior [1]. The prevalence of ASD increases globally, from 1 in 54 children in 2016 to 1 in 44 in 2018, and in Indonesia from 1 in 500 in 1995 to 1 in 50 in 2013, affecting an estimated 2.4 million individuals [2]. Children with ASD may not be fully independent, but they need to be able to perform daily activities, such as eating and drinking [3]. However, they often refuse food, become overly selective, or chew too slowly. Meal planning and support are essential to help them with their nutrition intake [4].

Prompts play a crucial role in helping children with ASD follow structured routines such as eating and drinking, and their consistent use across various settings enhances the effectiveness of daily activity support. Prompts can be physical, gestural, or verbal [5]. Prompts—whether verbal (e.g., "drinking") or gestural (e.g., pointing to a spoon)—serve as cues that guide the child through each step for a task. A child's response to prompts, such as picking up a spoon or bringing a cup to the mouth, indicates their understanding, attention, and task execution, providing valuable insight into functional performance in daily life.

Technologies in this field include activity prompt recognition, where image processing methods such as Histogram Equalization (HE), Contrast Stretching (CS), and Contrast Limited Adaptive Histogram Equalization (CLAHE) enhance image quality and reduce noise [6] and improve Convolutional Neural Network (CNN) performance [7].

Deep learning is effective for classification tasks due to its adaptable architecture and ability to automatically extract meaningful features from raw data [8]. The architecture of CNN can identify patterns within images by leveraging convolutional, pooling, and activation layers—collectively known as convolutional features [9].

To complement CNNs in recognizing motion information, Motion History Image (MHI), which is known for its computational efficiency and low memory usage [10]. MHI converts video sequences into static images. It is effective for motion representation as it can highlight recent object movements while minimizing background changes. In CNN,

MHI is a preprocessing step for motion recognition [11]. Numerous studies have employed MHI for action and activity recognition. MHI and Support Vector Machines (SVM) were used in a study by Tsai et al. [12] to recognize human actions. MHI and CNN were used in a study by Ahn et al. [13] to recognize cow actions. A study by Núñez et al. [14] integrated MHI with deep learning techniques to recognize daily activities, while Sahoo et al. [15] employed historical images along with CNN for activity recognition.

Previous studies [6, 16-19] employed CNN to identify ASD. The study by Singh et al. [16] demonstrated that CNN can be used to diagnose ASD from children's video recordings, achieving an accuracy of 85%. However, the study did not consider the impact of image or video quality on the model's performance, even though the dataset was collected from highly diverse YouTube sources with significant variations in lighting, noise, and resolution. Meanwhile, Haweel et al. [17] identified ASD using a CNN algorithm using brain image datasets, distinguishing between children with ASD and typically developing youngsters, and reached an 78% accuracy. However, this method remains limited to neuroimaging data characterized by low temporal resolution, high cost, and dependence on clinical facilities. Moreover, the proposed model primarily focuses on brain frequency features and does not incorporate directly observable behavioral aspects relevant to ASD detection. Furthermore, Sherkatghanad et al. [18] utilized open-access brain imaging datasets, and the experiments demonstrated that the CNN-based model achieved an accuracy of 70.22% in classifying individuals with ASD. Next, Huang [19] employed fMRI data and an SVM technique, attaining an accuracy of 70% in ASD classification. Both studies [18, 19] focused on diagnosing ASD in children using deep learning or machine learning approaches; however, these studies did not include feature extraction processes and were limited to neuroimaging data without considering behavioral aspects of ASD. In addition, previous work [6] on identifying daily activities in children with ASD included six drinking and eight eating sequences, achieving an accuracy of 85%. Although the performance was satisfactory, the study did not incorporate feature extraction, resulting in high computational complexity. This limitation makes such an approach less suitable for real-time applications that require efficient processing.

This study extends the work presented in reference [6] by focusing on detecting prompts within daily activities and integrating image enhancement with motion-based feature extraction using MHI within CNN architectures. Unlike previous studies that utilized uncontrolled online video sources, this work employs real-world activity videos recorded by therapists and parents under guided observation to ensure data reliability. The video was processed using three video enhancement techniques—HE, CS, and CLAHE—as applied in the reference [6]. Enhancement results were quantitatively evaluated to assess image quality improvement, and the best-performing method was integrated with MHI-based motion features for temporal representation before CNN-based prompt recognition. This integration addresses the visual variability and motion inconsistency commonly found in real-world ASD datasets, improving both recognition robustness and computational efficiency.

The proposed study introduces a novel integration of image enhancement and MHI within CNN architectures, namely VGG19 and MobileNetV2. This study contributes to literature in three key ways. First, the study employs real-world video performed by children with ASD, recorded by therapists and parents in natural, uncontrolled environments. Second, the study investigates image enhancement techniques to improve the recognition accuracy of these prompts under challenging visual conditions. Third, it proposes a method for classifying prompt types by applying motion-based feature extraction using MHI and image enhancement techniques, followed by activity recognition using a CNN across multiple datasets with enhanced image contrast. Together, these contributions highlight the study's novelty in bridging temporal and visual domains through image enhancement and motion representation for robust and efficient prompt recognition in children with ASD.

2. METHODOLOGY

The study is conducted in seven steps: collecting data, labeling videos, converting to grayscale, enhancing videos, extracting features, performing recognition, and evaluating results. The study stages are depicted in Figure 1.

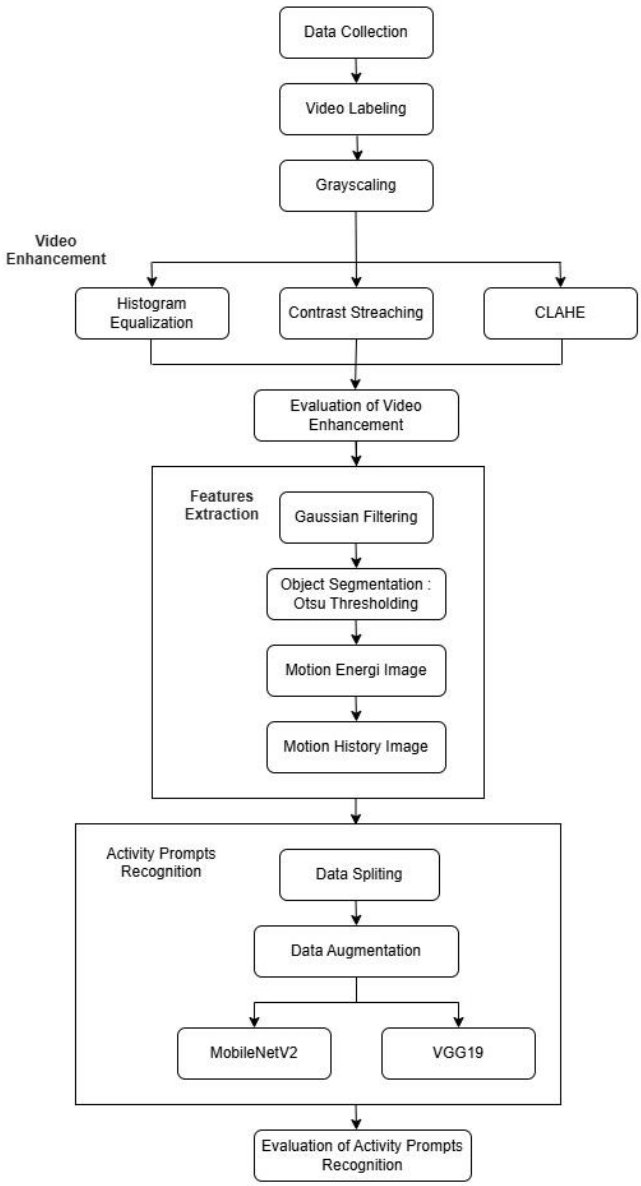


Figure 1. The study stages

2.1 Collecting data

The Ethics Committee of the Faculty of Public Health, Universitas Airlangga, reviewed and approved the data collection process for this study (clearance number: 76/EA/KEPK/2023). A total of 18 children aged 4 to 12 years participated in the study, consisting of eight students from the Public Special Needs School (SLB) Indonesia and ten from the Regional Technical Implementation Unit (UPTD) for children with Special Needs (ABK) Indonesia.

The collection of videos was performed after parents provided informed consent, which involved a detailed explanation of the study, followed by the completion and signing of consent forms by those who agreed to participate. The process was conducted on July 12–13, 2023. Data collection starts on July 17, 2023.

The video recording technique was carried out as follows: At UPTD ABK Indonesia, therapists and parents recorded activities during snack time, and at SLB Indonesia collected the data after school hours. At home, parents used mobile phones to record their children during breakfast or lunch. Each video captured either the full activity or selected segments, lasting between 2 and 5 minutes. Activities were recorded in portrait orientation with children seated on a chair or the floor, and the videos were later submitted to therapists or teachers.

2.2 Labeling videos

The recordings included entire sequences or specific segments of eating and drinking activities. Segmented videos were labeled individually, while full-length videos were manually divided and labeled according to therapist instructions. Eating activities had eight sequences, and drinking activities had six, each guided by specific prompts for the children.

According to Roncati et al. [5], prompts can be physical, gestural, or verbal. A physical prompt involves hand-over-hand guidance to assist with correct actions. A gestural prompt uses body movements, like pointing or hand gestures, to indicate the correct action. A verbal prompt provides spoken cues, including instructions, keywords, reminders, or questions [20].

(a) Physical (b) Gestural

(c) Verbal

Figure 2. Prompt types

Figure 2 illustrates three types of prompts. Figure 2(a) depicts a physical prompt, where the therapist helps through direct physical contact; this image was extracted from the file physical (35) at frame 24. Figure 2(b) represents a gestural prompt, in which the therapist offers guidance through body movements without physical touch; it was taken from the file gesture (36) at frame 35. Figure 2(c) illustrates a verbal prompt, where the therapist delivers instructions orally; this image was obtained from the file verbal (307) at frame 101.

The dataset contained 1,083 videos, with 606 eating videos and 477 drinking videos. The eating videos were divided into physical, gesture, and verbal, with respective counts of 159, 97, and 350. Similarly, drinking videos were split into physical, gesture, and verbal, with respective counts of 98, 75, and 304. Tables 1 and 2 provide details about the eating and drinking videos.

Table 1. Number of prompts in the eating sequence dataset

Activity Label	Number of Videos		
	Physical	Gesture	Verbal
Wash_hands	22	3	32
Grab_plate	11	2	15
Prepare_food	53	11	58
Take_packed_food	5	11	12
Open_packed_food	4	10	26
Pray	29	38	147
Eat	14	7	34
Finish_eating	21	15	26
Total	159	97	350

Table 2. Number of prompts in the drinking sequence dataset

Activity Label	Number of Videos		
	Physical	Gesture	Verbal
Grab_cup	18	25	59
Open_bottle	10	6	33
Pour_water	20	5	30
Drink	20	14	80
Close_bottle	9	5	34
Finish_drinking	21	20	68
Total	98	75	304

2.3 Converting to grayscale

Video frames were converted to gray scale to reduce computational complexity while maintaining recognition accuracy. A greyscale image shows different shades of grey based on pixel intensity, which is determined by bit depth. For example, an 8-bit image has 256 levels ranging from 0 to 255. In this study, video data were processed by extracting frames at a rate of 25–30 frames per second. converting each frame to grayscale, and applying a masking technique to preserve the anonymity of the children [21].

2.4 Enhancing videos

Image enhancement improves visual quality by refining contrast, sharpness, and clarity [22]. This study employs HE, CS, and CLAHE for video enhancement. HE redistributes pixel intensity for balanced contrast, CS increases intensity differences, and CLAHE enhances local contrast to reveal details in dark or bright areas.

Figure 3 illustrates the results of contrast enhancement. The original grayscale image is displayed in Figure 3(a), which features a black-to-white gradient representing the baseline

before enhancement. In Figure 3(b), the enhancement lightens the dark areas and darkens the light areas, resulting in increased contrast. More enhanced contrast, sharpened details, and clearer separation between light and dark regions are observed, as illustrated in Figure 3(c), while adaptively enhanced contrast that reveals obscured details is presented in Figure 3(d).

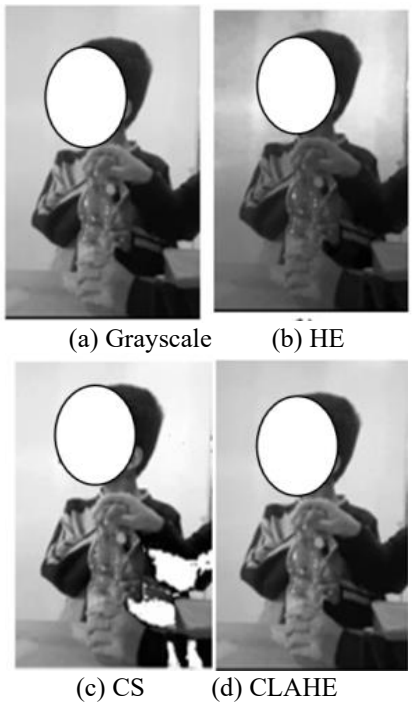


Figure 3. Result of image enhancement

2.5 Extracting features

Before feature extraction, Gaussian filtering was applied to the enhanced images obtained from the best-performing enhancement method. This process aimed to reduce high-frequency noise while preserving spatial smoothness and structural continuity across frames [23]. Gaussian filtering was chosen over median filtering because it is more suitable for real-world video data, effectively suppressing natural noise without distorting fine motion details or disrupting temporal consistency. This balanced smoothing maintains the integrity of subtle gestures and object edges, providing stable inputs for subsequent motion-based feature extraction such as MHI.

MHI was used for feature extraction due to its computational efficiency and low memory requirements. Unlike more complex temporal modeling approaches such as optical flow, 3D CNNs, or RNN-based architectures, MHI effectively captures recent motion cues through frame differencing while suppressing background information. This balance between accuracy and computational efficiency makes MHI particularly suitable for real-time applications.

Feature extraction occurs in three stages: object segmentation, MEI, and MHI. First, object segmentation is performed frame by frame using Otsu thresholding to separate image regions based on grayscale intensity. Otsu thresholding is a common segmentation technique that separates objects from the background based on pixel intensity while providing information on object size, position, and noise [24, 25].

Motion detection is deferred until the final stage, where segmentation is integrated with MEI and MHI to represent both object presence and movement. The objective is to separate objects from the background, creating binary images where objects are marked as 1 and the background as 0. After segmentation, MEI is extracted, representing motion-active regions with white pixels and static areas with black pixels.

We utilize only MEI and MHI. MHI at time t is calculated from MHI at time $t-1$ and the current motion image $D_t(x,y)$, whereas MEI is obtained by MHI thresholding. This recursive approach does not need to store or process previous image histories or their motion fields, making it computationally efficient and memory-saving. In comparison, various projection operators, such as pixel-wise summation over time, require retaining all $D_t(x,y)$ for $t_0 < t < t_\tau$.

Figure 4 shows a person sitting down. The top row represents critical frames from sitting sequences, while the bottom row displays cumulative binary motion pictures computed from the initial frame to the frame above. The cumulative binary motion images are known as MEI. Let $I(x,y,t)$ represent an image sequence and let $D(x,y,t)$ be a binary image sequence marking motion regions, where, in many applications, simple image differencing is sufficient to generate D. Thus, the binary MEI, denoted as $E_\tau(x,y,t)$, is defined in Eq. (1) [26].

$$E_\tau(x,y,t) = \cup_{i=0}^{\tau-1} D(x,y,t-i) \quad (1)$$

To extract the features of dynamic gestures and convert them into static images, MHI is employed as it displays the cumulative motion of objects along with a gradient trail. MHI is represented as $H_\tau(x,y,t)$, computed using Eq. (2) [10].

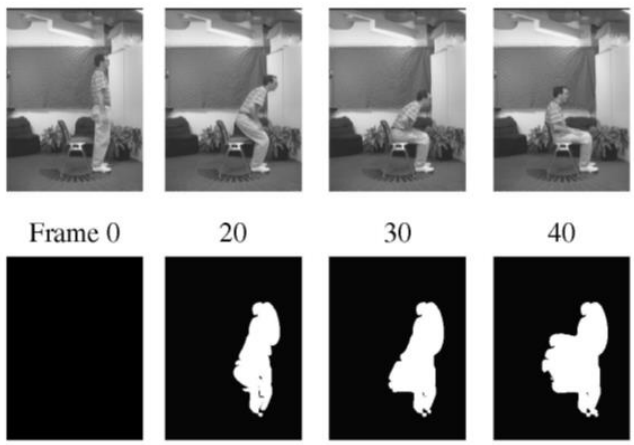


Figure 4. Sitting down keyframes (top) and cumulative motion images from Frame 0 (bottom) [26]

$$H_\tau(x,y,t) \begin{cases} \tau, & if\ \psi(x,y,t) = 1 \\ \max(0, H_\tau(x,y,t-1) - \delta, otherwise & \end{cases} \quad (2)$$

where, x and y indicate the position of pixels within the picture, and t denotes time. $\psi(x,y,t)$ is a binary image that records object motion in the present video frame and refreshes with each newly processed frame in the sequence. In this image, white pixels are assigned a value of 255, and black pixels 0. The period τ represents the timespan of a motion (measured in frames) and is usually defined by the total frame count of the video segment. A smaller value of τ will cause motion information in the MHI to decay too quickly, while a greater number may mask pixel value fluctuations (brightness variations) in the MHI. The decay parameter δ quantifies the

decrease in pixel values from prior frames as new frames are processed. If no new movement overlaps pixels previously occupied by motion in earlier frames, their values will decrease by δ, which is typically set to 1.

Frame subtraction is then performed to generate a binary image. If the discrepancies $D(x, y, t)$ between two consecutive frames surpass the threshold ξ, a binary image $\psi(x, y, t)$ is produced using Eq. (3).

$$\psi(x,y,t) = \begin{cases} 1 & if\ D(x,y,t) \geq \xi \\ 0 & otherwise \end{cases} \quad (3)$$

In this context, $\psi(x, y, t)$denotes the binary image at the th frame, while ξ is employed as a threshold to suppress background noise in the MHI. The frame discrepancy $D(x, y, t)$ is determined in Eq. (4).

$$D(x,y,t) = |I(x,y,t) - I(x,y,t - \triangle)| \quad (4)$$

where, $I(x, y, t)$ denotes the intensity at pixel coordinates (x, y) in the th frame of the image sequence, ranging from [0,255]. The symbol $\triangle$ symbolizes the temporal difference between two pixels at the same place. It is set to one to consider all frames.

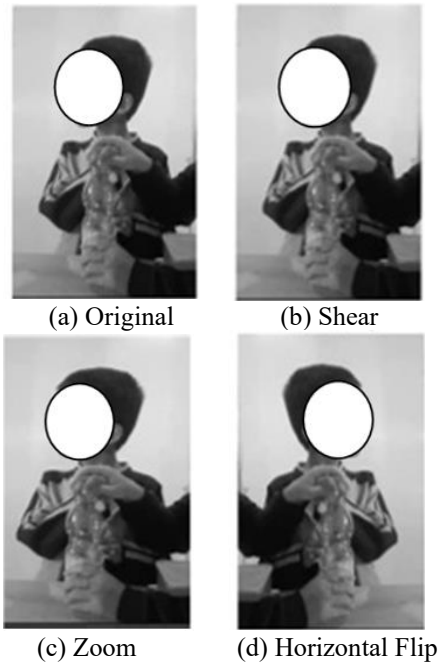
2.6 Performing recognition

After the feature extraction, the recognition process was performed using CNN based on the VGG19 and MobileNetV2 architectures. Before the recognition procedure, the dataset was randomly separated into two subsets: 80% training and 20% testing. Following splitting, data augmentation was applied only to the training set to increase variability and improve model generalization. Three augmentation techniques were used: zoom, shear, and horizontal flip. Specifically, 20% shearing was applied to introduce slanted transformations and increase viewpoint variation, 20% zooming was used to randomly scale the images for robustness to object size, and horizontal flipping was employed to introduce orientation diversity by randomly flipping the images left or right. Figure 5 illustrates the outcomes of the image augmentation process. Figure 5(a) shows the original image, while Figure 5(b) shows the effect of the shearing transformation, Figure 5(c) shows the result of the zoom transformation, and Figure 5(d) shows the output of the horizontal flip.

The next stage of model training utilized VGG19 and MobileNetV2. VGG19 is a deep CNN with 19 layers and 3 × 3 convolutional filters that were fine-tuned with a pre-trained ImageNet model [27] and resized input data of 224 × 224 pixels. MobileNetV2 is an improved version of MobileNetV1 optimized for resource-limited devices [28]. The VGG19 architecture employed in this study, as illustrated in Table 3, was generated from our Python-based implementation and used as the base training model, following Durai et al. [29].

Like VGG19, the pre-trained layers remained unchanged. The final convolutional layer used Global Average Pooling (GAP) to minimize feature dimensions, and the output was input into the classification layer using three neurons. MobileNetV2 architecture is presented in Table 4.

Table 5 presents eight proposed approaches: (1) MHI and VGG19 with CLAHE, (2) MHI and VGG19 with Grayscale, (3) MHI and MobileNetV2 with CLAHE, and (4) MHI and MobileNetV2 with grayscale. Four models for each of the eating and drinking datasets.



(a) Original (b) Shear

(c) Zoom (d) Horizontal Flip

Figure 5. Results of data augmentation

Table 3. VGG19 architecture based on our Python implementation, following the model by Durai et al. [29]

Layer (Type)	Output Shape	Param #
input_1 (InputLayer)	(None, 224, 224, 3)	0
block1_conv1 (Conv2D)	(None, 224, 224, 64)	1,792
block1_conv2 (Conv2D)	(None, 224, 224, 64)	36,928
block1_pool (MaxPooling2D)	(None, 112, 112, 64)	0
block2_conv1 (Conv2D)	(None, 112, 112, 128)	73,856
block2_conv2 (Conv2D)	(None, 112, 112, 128)	147,584
block2_pool (MaxPooling2D)	(None, 56, 56, 128)	0
block3_conv1 (Conv2D)	(None, 56, 56, 256)	295,168
block3_conv2 (Conv2D)	(None, 56, 56, 256)	590,080
block3_conv3 (Conv2D)	(None, 56, 56, 256)	590,080
block3_conv4 (Conv2D)	(None, 56, 56, 256)	590,080
block3_pool (MaxPooling2D)	(None, 28, 28, 256)	0
block4_conv1 (Conv2D)	(None, 28, 28, 512)	1,180,160
block4_conv2 (Conv2D)	(None, 28, 28, 512)	2,359,808
block4_conv3 (Conv2D)	(None, 28, 28, 512)	2,359,808
block4_conv4 (Conv2D)	(None, 28, 28, 512)	2,359,808
block4_pool (MaxPooling2D)	(None, 14, 14, 512)	0
block5_conv1 (Conv2D)	(None, 14, 14, 512)	2,359,808
block5_conv2 (Conv2D)	(None, 14, 14, 512)	2,359,808
block5_conv3 (Conv2D)	(None, 14, 14, 512)	2,359,808

block5_conv4 (Conv2D)	(None, 14, 14, 512)	2,359,808
block5_pool (MaxPooling2D)	(None, 7, 7, 512)	0
flatten (Flatten)	(None, 25088)	0
fc1 (Dense)	(None, 4096)	102,764,544
fc2 (Dense)	(None, 4096)	16,781,312
dense (Dense)	(None, 3)	12,291
Total params: 139,582,531		
Trainable params: 119,558,147		
Non-trainable params: 20,024,384		

Table 4. MobileNetV2 architecture based on our Python implementation

Layer (Type)	Output Shape	Param #
mobilenetv2_1.00_224_input (InputLayer)	[(None, 320, 320, 3)]	0
mobilenetv2_1.00_224 (Functional)	(None, 10, 10, 1280)	2,257,984
global_average_pooling2d_3 (GlobalAveragePooling2D)	(None, 1280)	0
fc1 (Dense)	(None, 4096)	5,246,976
fc2 (Dense)	(None, 4096)	16,781,312
predictions (Dense)	(None, 3)	12,291
Total params: 41,079,875		
Trainable params: 38,821,891		
Non-trainable params: 2,257,984		

Table 5. The proposed approaches

Proposed Approach	Dataset	Video Enhancement	Features Extraction	Model
1	Eating activities	CLAHE	MHI	VGG19
2		Without		
3		CLAHE		MobileNetV2
4		Without		
5	Drinking activities	CLAHE		VGG19
6		Without		
7		CLAHE		MobileNetV2
8		Without		

2.7 Evaluating results

Video enhancement evaluation using Means Square Error (MSE), Structural Similarity Index Measure (SSIM), and Peak Signal to Noise Ratio (PSNR). MSE is used to measure the similarity between reconstructed images or videos and their originals, as described in Eq. (5) [30].

$$MSE = \sum_{m=0}^{m} \sum_{n=0}^{n} ||(c(m,n) - s(m,n)|| \quad (5)$$

where, m and n denote the width and height of the cover image. The cover image is denoted by c, and the steganographic image after embedding as s.

The PSNR, measured in dB, evaluates the quality of processed images using pixel-wise comparisons to assess coding effectiveness, as shown in Eq. (6). The SSIM compares reconstructed image/video quality to its originals, as shown in Eq. (7).

$$PSNR = 10 . \log_{10} \frac{255^2}{\sqrt{MSE}} \quad (6)$$

$$SSIM = \frac{(2\mu_a\mu_b+C_1)+(2\sigma_{ab}+C_2)}{(\mu_a^2+\mu_b^2+C_1)+(\sigma_a^2+\sigma_b^2+C_1)} \quad (7)$$

In this case, a and μ_a indicate the original image and its mean, whereas b and μ_b denote the modified image and its mean. The covariance of both images is represented by σ_{ab}. The variables C_1 and C_2 are used to stabilise the division when the denominator is weak. The differences between the original and changed photographs are given as σ_a^2 and σ_b^2, respectively.

Here, a and μ_a denote the original image and its corresponding mean, whereas b and μ_b refer to the modified image and its mean. The covariance between the two images is expressed as σ_{ab}. To prevent instability when the denominator is close to zero, the constants C_1 and C_2 are introduced. The variances of the original and modified images are represented by σ_a^2 and σ_b^2, respectively.

Evaluation measures (accuracy, precision, recall, and F1 Score) are used to assess the efficiency of the technique, as shown in Eqs. (8)-(11). The performance of the CNN technique is further tested using a confusion matrix [8].

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (8)$$

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

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

$$F1\ Score = \frac{2\times Precision\times Recall}{Precision+Recall} \quad (11)$$

2.8 Area under Curve - Receiver Operating Characteristic (AUC-ROC)

The Receiver Operating Characteristic (ROC) curve assesses the effectiveness of classification models [31], which has a metric called the Area Under Curve (AUC), ranging between 0.5 and 1, indicating how effectively a test distinguishes between populations based on a specific condition. An AUC score of 0.5 reflects a test with no discriminative capability, equivalent to chance-level performance, whereas an AUC of 1.0 signifies ideal class separation [32]. The optimal point for an ROC curve is the upper left-hand corner, where the TPR is 1 and the FPR is 0, corresponding to an AUC-ROC of 1.

3. RESULTS AND DISCUSSION

This part describes the experimental findings of recognition (integration of CLAHE and MHI with CNN), compares them to state-of-the-art methods, and discusses.

3.1 Video enhancement evaluation

Video enhancement evaluation for eating activities is displayed in Tables 6-8. The average MSE for HE is 98.702, CS is 91.937, and CLAHE is 10.246. The average PSNR values are 17.431 for HE, 11.302 for CS, and 38.391 for CLAHE. The average SSIM is 0.860 for HE, 0.801 for CS, and 0.997 for CLAHE.

Table 6. The average MSE of eating activities

No.	Prompts	HE	CS	CLAHE
1	Physical	95.431	81.603	8.313
2	Gesture	98.855	100.625	11.570
3	Verbal	101.821	93.583	10.856
Average		98.702	91.937	**10.246**

Table 7. The average PSNR of eating activities

No.	Prompts	HE	CS	CLAHE
1	Physical	16.988	11.605	39.270
2	Gesture	18.321	11.103	37.748
3	Verbal	16.986	11.197	38.155
Average		17.431	11.302	**38.391**

Table 8. The average SSIM of eating activities

No.	Prompts	HE	CS	CLAHE
1	Physical	0.869	0.817	0.998
2	Gesture	0.871	0.799	0.997
3	Verbal	0.841	0.786	0.997
Average		0.860	0.860	**0.997**

Table 9. The average MSE of drinking activities

No.	Prompts	HE	CS	CLAHE
1	Physical	93.834	89.433	9.968
2	Gesture	97.468	94.987	11.303
3	Verbal	101.697	96.660	10.485
Average		97.667	93.693	**10.586**

Table 10. The average PSNR of drinking activities

No.	Prompts	HE	CS	CLAHE
1	Physical	17.475	11.285	38.635
2	Gesture	16.921	11.237	37.960
3	Verbal	16.675	11.249	38.229
Average		17.024	11.257	**38.275**

Table 11. The average SSIM of drinking activities

No.	Prompts	HE	CS	CLAHE
1	Physical	0.867	0.812	0.998
2	Gesture	0.834	0.781	0.997
3	Verbal	0.828	0.793	0.998
Average		0.843	0.796	**0.997**

Tables 9-11 display video enhancement evaluation for drinking activities. The average MSE result is 97.667 for HE, 93.693 for CS, and 10.586 for CLAHE. The average PSNR for HE is 17.024, CS is 11.257, and CLAHE is 38.275, while the average SSIM values are 0.843 for HE, 0.796 for CS, and 0.997 for CLAHE.

3.2 Extracting features

The feature extraction process was applied to a sample video from the eating activity dataset, specifically "Verbal (235).mp4," which consists of 124 extracted frames, as illustrated in Figure 6. A MHI representation is shown for frames 15 to 30. The frames were first enhanced using CLAHE, as depicted in Figure 6(a). Object segmentation was performed using Otsu thresholding, with the result shown in Figure 6(b). Subsequently, an MEI was generated in Figure 6(c), followed by the final MHI representation shown in Figure 6(d).

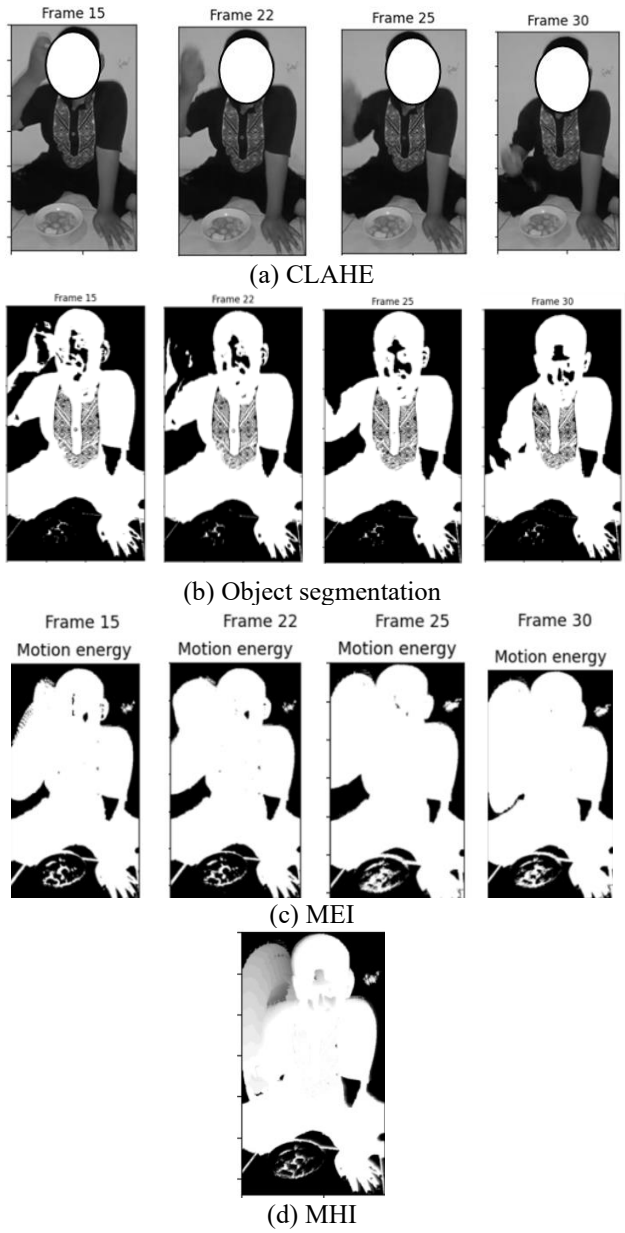


Figure 6. Result of feature extraction

3.3 Evaluation of activity prompts recognition

This study applies integration of CLAHE and MHI with VGG19 or MobileNetV2 to enhance recognition activity

prompts, specifically for eating and drinking, using evaluation metrics and AUC-ROC.

Table 12 presents the recognition result of the proposed approach. Proposed approach 1 achieved the highest accuracy of 71.31%, while proposed approach 4 yielded the lowest accuracy at 68.85%. Meanwhile, proposed approach 5 obtained the best performance with an accuracy of 73.96%, whereas proposed approach 8 recorded the lowest accuracy at 67.71%.

Table 12. Recognition results of the proposed approach for eating and drinking activities

Proposed Approach	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
1	71.31	70.12	71.31	67.70
2	69.67	59.25	69.67	62.70
3	70.49	74.08	70.49	69.26
4	68.85	73.39	68.85	63.39
5	73.96	73.51	73.96	73.70
6	72.92	71.02	72.92	71.05
7	69.79	69.07	69.79	63.08
8	67.71	66.16	67.71	60.70

Table 13. Computation time of the proposed method

Proposed Approach	Computation Time (seconds)
1	543
2	648
3	566
4	596
5	308
6	420
7	308
8	452

Table 13 presents the computation time of the proposed method. Proposed approach 1 achieved the highest accuracy with a computation time of 543 seconds, while proposed approach 4 yielded the lowest accuracy with a computation time of 596 seconds. Meanwhile, proposed approach 5 obtained the best performance with a computation time of 308 seconds, whereas proposed approach 8 recorded the lowest accuracy with a computation time of 452 seconds.

Based on Tables 12 and 13, MHI and (VGG19 and MobilenetV2) with enhancement (CLAHE) yield higher accuracy than MHI and (VGG-9 and MobileNetV2) without enhancement (grayscale) models. The use of CLAHE and MHI with (VGG19 and MobileNetV2) improves accuracy, indicating that CLAHE enhances the visibility of motion features in MHI, enabling more effective pattern recognition, while MHI and CLAHE reduce computation time.

Table 14 presents the AUC-ROC results of the eight proposed methods. The average AUC-ROC for the physical prompt is 0.82, for the gesture prompt 0.65, and the verbal prompt 0.75. The physical prompt has the highest average score, while the gesture prompt has the lowest. The fourth method achieves the highest AUC-ROC (0.94) for physical prompts, indicating strong classification ability with minimal misclassification risk. In contrast, the eighth method has the lowest AUC-ROC (0.60) for gesture prompts, showing difficulty in gesture classification.

Table 14. Results of AUC-ROC

Proposed Approach	The Prompting Method		
	Physical	Gesture	Verbal
1	0.92	0.77	0.82
2	0.92	0.77	0.82
3	0.92	0.62	0.81
4	0.94	0.69	0.8
5	0.86	0.88	0.85
6	0.87	0.85	0.85
7	0.82	0.65	0.75
8	0.83	0.60	0.76
Average	0.89	0.75	0.81

3.4 Comparison of proposed approaches

The suggested approach's conclusions are evaluated by comparing them to current approaches. Table 15 provides a comparison of the recommended techniques.

Table 15. Comparison of proposed approaches

Reference	Dataset	Baseline Architecture	Accuracy (%)	Computation Time (seconds)
Tsai et al. [12]	KTH human action	MHI and SVM	67.17	1.278
Ahn et al. [13]	Real Time cow action	MHI and SVM	72	-
Núñez et al. [14]	MSRDailyActivity3D	CNN	63.10	-
Sahoo et al. [15]	HMDB51	History Image and CNN	69.74	996
Huang et al. [19]		SVM	70	-
Sherkatghanad et al. [18]	fMRI	CNN	70.22	-
Haweel et al. [17]		CNN	78	-
Singh et al. [16]	Activity videos	MobileNetV1	85	-
Werdiningsih et al. [6]	Primary Data (eating and drinking data)	CLAHE + CNN	85	39.448
Proposed Approach - 1	Primary Data (eating data)	MHI and VGG19 with CLAHE	71.31	543
Proposed Approach - 2		MHI and VGG 19 with Grayscale	69.67	648
Proposed Approach - 3		MHI and MobilenetV2 with CLAHE	70.49	566
Proposed Approach - 4		MHI and MobilenetV2 with Grayscale	68.85	596
Proposed Approach - 5	Primary Data (drinking data)	MHI and VGG19 with CLAHE	73.96	308
Proposed Approach - 6		MHI and VGG 19 with Grayscale	72.92	420
Proposed Approach - 7		MHI and MobilenetV2 with CLAHE	69.79	308
Proposed Approach - 8		MHI and MobilenetV2 with Grayscale	67.71	452

Table 15 provides a comparative analysis of this study and previous research, emphasizing the impact of image enhancement on model performance. Previous studies [12-15] have shown that MHI combined with SVM achieved accuracies ranging from 67% to 72%, while CNN-based approaches for activity recognition reported accuracies between 63% and 70%. Nonetheless, the two methods have limitations in terms of precision and computational efficiency, with computation times in some situations exceeding 900 seconds. These findings underscore the need for approaches that improve both recognition performance and computational efficiency. In addition, several studies [16-19] investigated ASD diagnosis using secondary data obtained from public repositories. While most existing studies on ASD activity recognition report accuracy as the primary evaluation metric, they rarely address computational efficiency. In contrast, our study emphasizes real-time feasibility by reducing computation time while maintaining competitive accuracy.

Regarding comparison with prior studies in Table 15, several referenced works did not report computation time, limiting direct efficiency comparison. To ensure transparency, the current study explicitly reports computation time for each tested configuration, demonstrating that CLAHE integration not only improves accuracy but also reduces processing time, indicating practical gains in real-time applicability.

3.5 Discussion

Tables 6-11 reveal that CLAHE proves to be effective, as it achieves an average of 10.246 for MSE, 38.391 for PSNR, and 0.997 for SSIM in the eating dataset, as well as an average of 10.586 for MSE, 38.275 for PSNR, and 0.997 for SSIM in the drinking dataset. An MSE value near 0 and <30 indicates good results and reduced error. PSNR reflects image processing quality, with values >35 dB indicating high accuracy and <35 dB, suggesting otherwise. SSIM ranges from -1 to 1, where 1 signifies identical images, and a value near zero or negative reflects negligible similarity [33]. Considering these results, CLAHE was chosen as the image enhancement method for recognition.

Tables 12 and 13 show that CLAHE preprocessing improves classification performance compared to grayscale, increasing accuracy from 69.67% to 71.31% for eating activities and from 72.92% to 73.96% for drinking activities. This improvement is attributed to CLAHE's ability to enhance local contrast, which is particularly useful under varied lighting and background conditions common in real-world ASD recordings. Moreover, CLAHE integration reduced computation time, from 648 to 543 seconds for eating activities and from 420 to 308 seconds for drinking activities, demonstrating its dual benefit of improving accuracy and efficiency [34]. Although the numerical improvements in accuracy are modest (around 1–2%), the consistent gains across all metrics and the substantial reduction in computation time demonstrate the practical advantage of integrating CLAHE and MHI with CNN for real-time recognition tasks. These results indicate that between contrast optimization (CLAHE) and temporal motion encoding (MHI) enhances not only improve image quality but also facilitate faster convergence and more efficient feature extraction in CNN-based models.

In this study, each configuration was trained and tested multiple times, and the best-performing result was reported to reflect the model's optimal capability under each enhancement setting. Formal statistical significance testing (e.g., p-value computation) was not conducted because the evaluation focused on best-performing results rather than averaged outcomes across repeated runs. Future work will include statistical validation through multiple-run averaging to strengthen empirical reliability.

Table 14 presents the AUC-ROC results of the proposed approach, with the physical prompt achieving a score of 0.89, indicating strong classification performance. This suggests the model effectively distinguishes classes with minimal errors [31]. The AUC-ROC for the gesture prompt is the lowest at 0.75, indicating adequate performance but a higher likelihood of errors [32]. Gesture prompting is the least effective due to its ambiguity compared to physical and verbal methods. Physical prompts offer clear, direct cues, while verbal prompts provide explicit linguistic meaning, which is easier to interpret. Additionally, labeling gesture data is challenging, as it requires visual movement interpretation, increasing the risk of errors that can impact model accuracy.

Table 15 shows that applying video enhancement improves CNN classification accuracy. For the eating activities dataset, the MHI–VGG19 model with CLAHE achieves 71.31%, compared to 69.67% without enhancement (grayscale). For drinking activities, the same model reaches 73.96% with CLAHE, compared to 72.92% without enhancement. This improvement can be attributed to CLAHE, which increases image contrast and clarity, thereby enabling CNN to extract more discriminative features for classification. In addition, VGG19 demonstrates higher accuracy than MobileNetV2. Specifically, for eating activities, MHI–VGG19 with CLAHE achieves 71.31% while MHI–MobileNetV2 with CLAHE achieves 70.49%. For drinking activities, MHI–VGG19 with CLAHE achieves 73.96% compared to 69.79% for MHI–MobileNetV2. The superior performance of VGG19 can be explained by its deeper 19-layer architecture [27], which allows the extraction of more complex features and better detection of subtle motion patterns in MHI.

The relatively low classification accuracy in this study is primarily due to the dataset's inherent complexity and variability, like the Real-Time Cow Action [13] and HMDB51 [15] datasets, the data used in our experiments exhibit substantial variations in motion patterns, camera angles, and background clutter, which pose significant challenges for action recognition models. In addition, many gestures exhibited by children with autism tend to be subtle, non-standardized, and vary significantly across individuals, making them difficult to interpret—even for trained human observers [35]. These factors collectively represent a major challenge in developing robust image-based activity recognition systems for this population.

These characteristics restrict the model's ability to generalize across samples, lowering overall performance despite the robustness of the proposed strategy. Furthermore, the usage of MHI as the dominant temporal representation may add to the reduced accuracy, as MHI tends to oversimplify motion dynamics in complex scenes with overlapping actions, background motion, or occlusions. This limitation makes it difficult to capture fine-grained temporal cues required for distinguishing subtle action differences in highly variable datasets. Similar challenges in applying MHI to complex datasets have also been reported in previous studies, where the method showed limited effectiveness in capturing temporal nuances in unconstrained video scenarios [13, 15].

Table 15 shows that the accuracy achieved in this study is

lower than that reported in the study [6]. However, it reduces the computational time of from 39.448 seconds to 308 seconds. Notably, this shorter processing time corresponds to the highest accuracy obtained in our experiments, reaching 73.96%. This substantial improvement in computational efficiency enhances the system's feasibility for real-time detection scenarios, which is crucial in practical applications. Fast processing time directly contributes to better scalability [36], enabling the system to provide prompt responses during daily routines for children with ASD. By reducing the number of processed frames to one per video using MHI, the computational cost is minimized. Although this results in a loss of temporal information and may have a modest impact on classification accuracy, the trade-off is acceptable in the real world, resource-constrained environments where real-time interaction is essential.

Enhancing MHI with preprocessing methods (e.g., Gamma Correction [37], Gaussian Blur [38] can preserve motion patterns, while leveraging diverse datasets, pose estimation, and optical flow may reduce gesture ambiguity. Integrating CNN and RNN is expected to further improve recognition accuracy, reliability, and overall system performance.

4. CONCLUSIONS

This study introduced CNN-based approach for prompt activity recognition of children with ASD by integrating CLAHE and MHI with CNN architectures (VGG19 and MobileNetV2). The integration enhanced recognition robustness and computational efficiency, demonstrating that combining visual enhancement and temporal motion encoding supports accurate and real-time performance. Physical prompts were recognized with high confidence (AUC-ROC = 0.94), while gesture prompts remained more challenging (AUC-ROC = 0.60), highlighting the need for improved temporal feature modelling.

The relatively modest classification accuracy reflects the dataset's inherent complexity and variability. Variations in motion patterns, camera angles, and background clutter, along with subtle and inconsistent gestures among children with ASD, pose recognition challenges even for human observers. Moreover, this study did not include formal statistical validation, as the reported results represent best-performing outcomes rather than averaged experiments. These factors collectively represent key limitations in achieving consistent and statistically verified model performance.

Future work will address these limitations by incorporating multimodal learning (e.g., depth, skeletal, and audio cues), attention-based temporal modeling, and statistical validation through multiple-run averaging to strengthen empirical reliability and improve generalization across diverse real-world activities.

REFERENCES

[1] Genovese, A., Butler, M.G. (2020). Clinical assessment, genetics, and treatment approaches in autism spectrum disorder (ASD). International Journal of Molecular Sciences, 21(13): 4726. https://doi.org/10.3390/ijms21134726

[2] Daulay, N., Daulay, H., Rohman, F. (2025). Religious coping of Muslim mothers of children with autism spectrum disorder in Indonesia. Journal of Disability & Religion, 29(1): 33-50. https://doi.org/10.1080/23312521.2024.2372021

[3] Levy-Dayan, H., Josman, N., Rosenblum, S. (2023). Basic activity of daily living evaluation of children with autism spectrum disorder: Do-eat washy adaption preliminary psychometric characteristics. Children, 10(3): 514. https://doi.org/10.3390/children10030514

[4] Kazek, B., Brzóska, A., Paprocka, J., Iwanicki, T., Kozioł, K., Kapinos-Gorczyca, A., Likus, W., Ferlewicz, M., Babraj, A., Buczek, A., Krupka-Matuszczyk, I., Emich-Widera, E. (2021). Eating behaviors of children with autism—Pilot study, Part II. Nutrients, 13(11): 3850. https://doi.org/10.3390/nu13113850

[5] Roncati, A.L., Souza, A.C., Miguel, C.F. (2019). Exposure to a specific prompt topography predicts its relative efficiency when teaching intraverbal behavior to children with autism spectrum disorder. Journal of Applied Behavior Analysis, 52(3): 739-745. https://doi.org/10.1002/jaba.568

[6] Werdiningsih, I., Puspitasari, I., Hendradi, R. (2025). Recognizing daily activities of children with autism spectrum disorder using convolutional neural network based on image enhancement. Cybernetics and Information Technologies, 25(1): 78-96. https://doi.org/10.2478/cait-2025-0005

[7] Ferdinand, V.A., Henry, G.E. (2022). Effect of image enhancement in CNN-based medical image classification: A systematic literature review. In International Conference on Information and Communications Technology (ICOIACT), Yogyakarta, Indonesia, pp. 87-92. https://doi.org/10.1109/ICOIACT55506.2022.9972030

[8] Olayiwola, J.O., Badejo, J.A., Okokpujie, K., Awomoyi, M.E. (2023). Lung-related diseases classification using deep convolutional neural network. Mathematical Modelling of Engineering Problems, 10(4): 1097-1104. https://doi.org/10.18280/mmep.100401

[9] Yi, L., Siedler, C., Kinkel, Y., Glatt, M., Kölsch, P., Aurich, J. (2021). Object detection in factory based on deep learning approach. Procedia CIRP, 104: 1029-1034. https://doi.org/10.1016/j.procir.2021.11.173

[10] Koh, Y., Kim, T., Hong, M., Choi, Y.J. (2020). CNN-based gesture recognition using motion history image. Journal of Internet Computing & Services, 21(5): 67-73. https://doi.org/10.7472/jksii.2020.21.5.67

[11] Chen, H., Leu, M.C., Yin, Z. (2022). Real-time multi-modal human–robot collaboration using gestures and speech. Journal of Manufacturing Science and Engineering, 144(10): 101007. https://doi.org/10.1115/1.4054297

[12] Tsai, D.M., Chiu, W.Y., Lee, M.H. (2015). Optical flow-motion history image (OF-MHI) for action recognition. Signal, Image and Video Processing, 9(8): 1897-1906. https://doi.org/10.1007/s11760-014-0677-9

[13] Ahn, S.J., Ko, D.M., Heo, E.J., Choi, K.S. (2018). Real-time cow action recognition based on motion history image feature. In 2018 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, USA, pp. 1-2. https://doi.org/10.1109/ICCE.2018.8326090

[14] Núñez, J.C., Cabido, R., Pantrigo, J.J., Montemayor, A.S., Vélez, J.F. (2018). Convolutional neural networks and long short-term memory for skeleton-based human activity and hand gesture recognition. Pattern

Recognition, 76: 80-94. https://doi.org/10.1016/j.patcog.2017.10.033
[15] Sahoo, S.P., Ari, S., Mahapatra, K., Mohanty, S.P. (2020). HAR-depth: A novel framework for human action recognition using sequential learning and depth estimated history images. IEEE Transactions on Emerging Topics in Computational Intelligence, 5(5): 813-825. https://doi.org/10.1109/TETCI.2020.3014367
[16] Singh, A., Rawat, A., Laroia, M., Seeja, K.R. (2024). Autism spectrum disorder screening on home videos using deep learning, International Journal of Image, Graphics and Signal Processing (IJIGSP), 16(4): 106-115. https://doi.org/10.5815/ijigsp.2024.04.08
[17] Haweel, R., Shalaby, A., Mahmoud, A., Seada, N., Ghoniemy, S., Ghazal, M., El-Baz, A. (2021). A robust DWT-CNN-based CAD system for early diagnosis of autism using task-based fMRI. Medical Physics, 48(5): 2315-2326. https://doi.org/10.1002/mp.14692
[18] Sherkatghanad, Z., Akhondzadeh, M., Salari, S., Zomorodi-Moghadam, M., Abdar, M., Acharya, U.R., Khosrowabadi, R., Salari, V. (2020). Automated detection of autism spectrum disorder using a convolutional neural network. Frontiers in Neuroscience, 13: 1325. https://doi.org/10.3389/fnins.2019.01325
[19] Huang, B. (2019). Diagnosis of autism spectrum disorder by causal influence strength learned from resting-state fMRI data. Imaging and Signal Analysis Journal, 1: 237-267. https://doi.org/10.1016/B978-0-12-822822-7.00012-0
[20] Seaver, J.L., Bourret, J.C. (2014). An evaluation of response prompts for teaching behavior chains. Journal of Applied Behavior Analysis, 47(4): 777-792. https://doi.org/10.1002/jaba.159
[21] Yeh, C.H., Lin, C.H., Kang, L.W., Huang, C.H., Lin, M.H., Chang, C.Y., Wang, C.C. (2022). Lightweight deep neural network for joint learning of underwater object detection and color conversion. IEEE Transactions on Neural Networks and Learning Systems, 33(11): 6129-6143. https://doi.org/10.1109/TNNLS.2021.3072414
[22] Qi, Y., Yang, Z., Sun, W., Lou, M., Lian, J., Zhao, W., Ma, Y. (2022). A comprehensive overview of image enhancement techniques. Archives of Computational Methods in Engineering, 29(1): 583-607. https://doi.org/10.1007/s11831-021-09587-6
[23] Hemamalini, V., Rajarajeswari, S., Nachiyappan, S., Sambath, M., Devi, T., Singh, B.K., Raghuvanshi, A. (2022). Food quality inspection and grading using efficient image segmentation and machine learning-based system. Journal of Food Quality, 2022(1): 5262294. https://doi.org/10.1155/2022/5262294
[24] Nyo, M.T., Mebarek-Oudina, F., Hlaing, S.S., Khan, N.A. (2022). Otsu's thresholding technique for MRI image brain tumor segmentation. Multimedia Tools and Applications, 81(30): 43837-43849. https://doi.org/10.1007/s11042-022-13215-1
[25] Goh, T.Y., Basah, S.N., Yazid, H., Safar, M.J.A., Saad, F.S.A. (2018). Performance analysis of image thresholding: Otsu technique. Measurement, 114: 298-307. https://doi.org/10.1016/j.measurement.2017.09.052
[26] Bobick, A.F., Davis, J.W. (2002). The recognition of human movement using temporal templates. IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(3): 257-267. https://doi.org/10.1109/34.910878
[27] Killi, C.B.R., Balakrishnan, N., Rao, C.S. (2023). Deep fake image classification using VGG-19 model. Ingénierie des Systèmes d'Information, 28(2): 509-515. https://doi.org/10.18280/isi.280228
[28] Indraswari, R., Rokhana, R., Herulambang, W. (2022). Melanoma image classification based on MobileNetV2 network. Procedia Computer Science, 197: 198-207. https://doi.org/10.1016/j.procs.2021.12.132
[29] Durai, L., Daniel, J., Mathan, K. (2024). Generating artistic images through neural style transfer using machine learning and image manipulation techniques. International Journal of Scientific Research in Engineering and Management (IJSREM), 8(10): 1-30. https://doi.org/10.55041/IJSREM37820
[30] Archana, R., Jeevaraj, P.E. (2024). Deep learning models for digital image processing: A review. Artificial Intelligence Review, 57(1): 11. https://doi.org/10.1007/s10462-023-10631-z
[31] Khan, S.A., Rana, Z.A. (2019). Evaluating performance of software defect prediction models using area under precision-recall curve (AUC-PR). In International Conference on Advancements in Computational Sciences (ICACS), Lahore, Pakistan, pp. 1-6. https://doi.org/10.23919/ICACS.2019.8689135
[32] Hoo, Z.H., Candlish, J., Teare, D. (2017). What is an ROC curve? Emergency Medicine Journal, 34(6): 357-359. https://doi.org/10.1136/emermed-2017-206735
[33] Erwin, Ningsih, D.R. (2021). Improving retinal image quality using the contrast stretching, histogram equalization, and CLAHE methods with median filters. International Journal of Image, Graphics and Signal Processing, 14(2): 30-41. https://doi.org/10.5815/ijigsp.2020.02.04
[34] Alpan, K., Arman, B., Dimililer, K. (2025). Effect of Contrast Limited Adaptive Histogram Equalization (CLAHE) on breast cancer detection using residual network (ResNet). In International Conference on Computational Intelligence Approaches and Applications (ICCIAA), Amman, Jordan, pp. 1-5. https://doi.org/10.1109/ICCIAA65327.2025.11013776
[35] Logrieco, M.G., Annechini, E., Casula, L., Guerrera, S., Fasolo, M., Vicari, S., Valeri, G. (2024). Nonverbal skills evolution in children with autism spectrum disorder one-year post-diagnosis. Children, 11(12): 1520. https://doi.org/10.3390/children11121520
[36] Petricia, V.A., Jacksy, M., Bhavadharani, R. (2025). Autism spectrum disorder detection using convolutional neural network. In International Conference on Visual Analytics and Data Visualization (ICVADV), Tirunelveli, India, pp. 743-748. https://doi.org/10.1109/ICVADV63329.2025.10961573
[37] Jeon, J.J., Park, J.Y., Eom, I.K. (2024). Low-light image enhancement using gamma correction prior in mixed color spaces. Pattern Recognition, 146: 110001. https://doi.org/10.1016/j.patcog.2023.110001
[38] Devi, T.G., Patil, N., Rai, S., Philipose, C.S. (2023). Gaussian blurring technique for detecting and classifying acute lymphoblastic leukemia cancer cells from microscopic biopsy images. Life, 13(2): 348. https://doi.org/10.3390/life13020348