BULGARIAN ACADEMY OF SCIENCES

CYBERNETICS AND INFORMATION TECHNOLOGIES • Volume **25**, No 1
Sofia • 2025 Print ISSN: 1311-9702; Online ISSN: 1314-4081
DOI: 10.2478/cait-2025-0005

Recognizing Daily Activities of Children with Autism Spectrum Disorder Using Convolutional Neural Network Based on Image Enhancement

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, Indonesia*
[2]*Information Systems Study Program, Faculty of Science and Technology, Universitas Airlangga, Surabaya, Indonesia*
[3]*Research Center for Quantum Engineering Design, Faculty of Science and Technology, Universitas Airlangga, Surabaya, Indonesia*
*E-mails: indah-w@fst.unair.ac.id ira-p@fst.unair.ac.*id (corresponding author) *rimuljohendradi@fst.unair.ac.id*

Abstract: *Independence for individuals with disabilities, Children with Autism Spectrum Disorder* (*ASD*)*, need skills to perform daily activities. This study focuses on recognizing the daily activities of children with ASD using a Convolutional Neural Network* (*CNN*) *based on augmented images. The CNN architectures employed are Visual Geometry Group 19* (*VGG19*) *and MobileNetV2, while image improvement techniques include Histogram Equalization, Contrast Stretching, and Contrast Limited Adaptive Histogram Equalization* (*CLAHE*)*. The data consists of eating* (*606 videos*) *and drinking* (*477 videos*) *activities recorded by therapists or parents. CLAHE proved the most effective, achieving an SSIM of 0.998 and a PSNR of 38.466 for the eating activities, an SSIM of 0.998, and a PSNR of 38.296 for the drinking activities. Experimental results using CLAHE and VGG19 showed a recognition model accuracy of 85%, while VGG19 without image enhancement achieved an accuracy of 83%. CNN with image enhancement achieves slightly better accuracy, though the difference is insignificant.*

Keywords: *Autism, CNN, Disabilities, Image enhancement, Recognition.*

1. Introduction

Autism Spectrum Disorder (ASD), as described in the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) and International Classification of Diseases 11th revision (ICD11), is categorized as a “neurodevelopmental disorder”, characterized by deficits in cognition, communication, behavior, and/or motor skills resulting from atypical brain development. These disorders are predominantly marked by their early onset in childhood. They are associated with difficulties in personal, social, educational, and occupational development, with a tendency to co-occur with other conditions[1]. Developmental disorders are more accurately referred to as neurodevelopmental disorders. These conditions rooted in neurological differences can disrupt the development, retention, or use of skills or knowledge. They may affect areas such as attention, memory, perception, language, problem-solving, or social interaction. The severity of these disorders can vary, with some being mild and manageable through educational and behavioral strategies, while others may be more severe, requiring additional support for the affected individuals [2].

In Indonesia, cases of children with ASD have not yet received a positive response from the community. The existence of a negative stigma reveals that children with ASD are still seen as a family disgrace that must be hidden from the community. To reduce this negative stigma, autistic children need to develop the ability to live independently [3]. Living independently does not mean that people with disabilities are supposed to do everything on their own and do not need others, nor does it mean a desire to live alone and limit themselves within the confines of their family. It means they can also attend school, take public transportation, work, and even start their own family [4]. Therefore, certain skills are required to support independent living, including the ability to handle daily activities [5], such as eating and drinking [6].

Eating and drinking are primary needs, serving as a source of energy for all activities and one of the most important factors for survival. Children with ASD tend to have irregular eating habits and follow their mood or appetite. They often refuse to eat, are picky with food, are reluctant to try new foods, may exhibit tantrums, and chew very slowly. Children with ASD may have unusual eating patterns and behaviors [7]. Therefore, managing and preparing meals, as well as scheduling mealtimes, is an important program for children with ASD [8].

To monitor the development of children with ASD, educators or therapists often ask parents to send videos of their children’s eating and drinking activities at home. However, these videos often have issues, such as poor lighting quality and unclear movements, making it difficult for educators to recognize the activities. These videos can be managed with Human Activity Recognition (HAR), which identifies and analyzes human activities through inputs such as video or sensors [9]. These activities include walking, running, and exercising [10]. HAR is utilized in the healthcare field to monitor the activities of patients undergoing therapy [11].

Image processing is essential in preparing for HAR. The initial phase involves enhancing image quality to reduce noise and improve the accuracy of HAR outcomes. Images may contain a wealth of information, but their quality can often be degraded or flawed by noise. Various image processing techniques can be applied to transform images into higher-quality versions to facilitate image interpretation.

Convolutional Neural Networks (CNN) are used in HAR [12, 13]. Their main advantage is the ability to detect important features [14]. CNN architecture can recognize patterns in an image by utilizing convolution, pooling, and activation layers called convolutional features. Convolutional neural networks are artificial neural networks capable of tackling many computer vision tasks, such as image categorization, object identification, and general recognition [15]. A study by [16] aimed to diagnose ASD using CNN algorithms. The dataset used was open-source brain image data, specifically functional Magnetic Resonance Imaging (fMRI) data, classified as ASD or Typically Developing (TD). The results indicated that feature extraction combined with CNN could diagnose ASD with an accuracy of 78 %. Likewise, a study by [17] aimed to detect ASD using brain image datasets with CNN algorithms. The brain image dataset used was open-source fMRI data represented by the Autism Brain Imaging Data Exchange (ABIDE). The results showed that the CNN model could detect ASD with an accuracy of 70.22%.

The appropriateness of the image enhancement technique selected for CNN can positively impact performance [18]. Image enhancement seeks to increase the quality of information in images so that it may be read more properly, thereby boosting the dataset's classification performance [19]. As one implementation of image processing [20], image enhancement manipulates and adjusts images to make them more suitable for analysis.

A study by [21] aimed to enhance video image quality by improving contrast through the application of Histogram Equalization (HE), Contrast Stretching (CS), and Contrast Limited Adaptive Histogram Equalization (CLAHE) techniques. The study primarily concentrated on comparing different image enhancement techniques. The study utilized data on the eating activities of children with ASD. This research proposes several approaches, which are divided into four main steps or procedures. The first step involves data acquisition, where data is collected and labeled. Once the data collection is complete, the original RGB images are introduced. RGB images are composed of three primary colors: red, green, and blue. Next, these images are converted into grayscale, representing various gray levels. Following this, the video undergoes enhancement through triple contrast enhancement techniques, including HE, CS, and CLAHE. Finally, the evaluation phase is conducted, utilizing Mean Square Error (MSE) and Peak Signal Noise Ratio (PSNR).

The research to be carried out continues the research conducted in [21], with the addition of drinking data alongside the previously used eating data. This study consists of six steps: data collection, video labeling, gray scaling, video enhancement, video recognition, and evaluation. The eating and drinking data were processed using the same image enhancement methods as in [21], namely HE, CS, and CLAHE. The enhancement process was evaluated using Structural Similarity Index Measure (SSIM) and PSNR metrics. The enhancement method that yields the best results will be used to recognize the daily activities of children with ASD. CNN is employed as the model for recognizing these daily activities.

This study contributes to the literature in three ways. Firstly, data collected in this study consists of videos of children with ASD’s daily activities, such as eating and drinking, which were recorded directly by therapists or parents. The video captured the eating and drinking activities of children with ASD. Eating involves a sequence of activities: washing hands, using a plate and a spoon, putting the rice onto the plate, adding side dishes, praying, eating, and cleaning up after oneself. Likewise, drinking involves a sequence of activities: Taking a cup or a glass, pouring water, opening and closing a bottle, and drinking. The second contribution of this study is selecting the optimal image enhancement strategy for various datasets to improve the recognition of children with ASD's performance in completing daily activities. The video was extracted into several frames and processed using image processing methods. This process produces a video that is better than the original because the image quality of the video has been improved. The third contribution is to recognize children with ASD's daily activities using CNN architecture, which has been successful in various datasets with improved video quality (enhanced contrast).

2. Proposed methodology

This study consists of six steps: data collection, video labeling, gray scaling, video enhancement, video recognition, and evaluation. The videos were enhanced by HE, CS, and CLAHE. For video recognition, the Visual Geometry Group 19 (VGG19) and MobileNetV2 architectures were employed. The evaluations were conducted in two parts: video enhancement, which is assessed using SSIM and PSNR metrics, and video recognition, which is assessed using accuracy, recall, precision, and F1 score. The stages are illustrated in Fig 1.

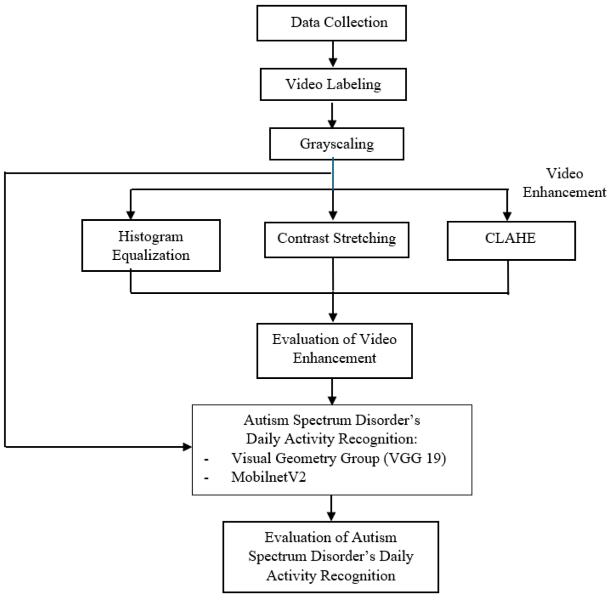

Data Collection
Video Labeling
Grayscaling
Video Enhancement
Histogram Equalization
Contrast Stretching
CLAHE
Evaluation of Video Enhancement
Autism Spectrum Disorder's Daily Activity Recognition:
- Visual Geometry Group (VGG 19)
- MobilnetV2
Evaluation of Autism Spectrum Disorder's Daily Activity Recognition

Fig. 1. The study stages

2.1. Data Collection

Data collection This study included four stages: obtaining an ethical clearance letter, securing research permissions, obtaining informed consent, and data collection. The FKM Universitas Airlangga ethics team has reviewed this study and passed the ethical clearance and review process, issuing an ethical clearance letter with the number 76/EA/KEPK/2023. The study has also received approvals from the faculty, National Unity and Politics Agency or Badan Kesatuan Bangsa dan Politik (Bakesbangpol) East Java Province, Bakesbangpol Sidoarjo, Department of Education and Culture of Sidoarjo, and Regional Technical Implementation Unit or

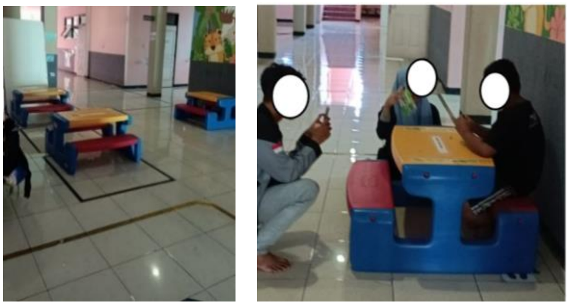
Unit Pelaksana Teknis Daerah (UPTD) of Children with Special Needs or Anak Berkebutuhan Khusus (ABK) Sidoarjo. Informed consent was obtained by explaining it to parents, after which those who agreed filled out and signed the informed consent form. This process took place on July 15-16, 2023. The videos were collected after obtaining informed consent, starting on July 17, 2023. Data collection was carried out over a period of 6 months.

Video recordings of the eating and drinking activities of children with ASD are used as primary data in this investigation. The video recording process is illustrated in Fig 2. Data was collected from:

a. UPTD ABK Sidoarjo, located at Mendalan Street IV No 8, Regency of Sidoarjo.

b. Public Special Needs School or Sekolah Luar Biasa (SLB) Negeri Lamongan, located at Mendalan Street No 6, Regency of Lamongan.

c. Ma’arif Nu Private Special Needs School Lamongan, located at Village of Bakalanpule, District Tikung, Regency of Lamongan.



(a) (b)

Fig. 2. The process of video recording: The space for snack time (a); The video recording (b)

The study involved 18 students, eight from SLB Lamongan and ten from UPTD ABK Sidoarjo, aged between 4 and 12 years. The video recording process followed this procedure in order:

1. Videos at UPTD ABK Sidoarjo were recorded by the therapists and parents.
2. Videos at SLB Lamongan were recorded by the teacher.
3. Videos at UPTD ABK Sidoarjo were recorded during snack time.
4. Videos at SLB Lamongan were recorded after school.
5. Videos of eating and drinking activities were recorded by parents at home using their mobile phones.
6. Videos were recorded by parents when the child was having breakfast or lunch.
7. Videos should capture the entire activity or its segments, lasting 2-5 minutes.
8. Videos were recorded in portrait mode (vertical) to capture the eating and drinking activities optimally.
9. Children were sitting on a chair or the floor.
10. Videos recorded by parents were sent to the therapists or the teacher.

2.2. Video labeling

The video recordings were either segmental sequences or complete videos capturing the entire eating and drinking activities from start to finish. If the submitted videos were segmented, each segment was labeled. Conversely, if the video covered the entire eating and drinking process, it was manually divided into several segments, each representing a sequence of eating and drinking activities. Subsequently, each segment was labeled, such as taking a plate, preparing the food, eating, etc.

Eating activities were tagged with eight labels and drinking activities with six labels. There were 1083 videos in the dataset, divided into 606 eating activity videos and 477 drinking activity videos. Tables 1 and 2 present information about the videos of the eating and drinking activities.

Table 1. Information about the eating activity videos

Activity label	Number of videos	Average video duration (s)
Washing hands	57	4.82
Taking a plate	28	3.39
Preparing the food	121	10.83
Taking packed food	28	3.89
Opening packed food	40	6.45
Praying	55	23.04
Eating	215	9.77
Finishing eating	62	5.34
Total	606	

Table 2. Video information on drinking activities

Activity label	Number of videos	Average video duration (s)
Taking a drinking cup	102	2.95
Opening the bottle	49	3.37
Pouring the water	55	5.4
Drinking	114	6.11
Closing the bottle	48	4.67
Finishing drinking	109	2.23
Total	477	

2.3. Gray scaling

A grayscale image is defined by pixel intensity values representing different shades of gray. These values are limited by the minimum and maximum grey levels, which depend on the bit-depth used. For example, an 8-bit image provides 256 levels of gray, with 0 being the darkest and 255 being the brightest [22].

Each video was extracted into frames, with frames taken between 25-30 frames per second (fps) according to the original fps in the video. Each frame extracted from the video was converted to grayscale. The images were intentionally masked to avoid disclosing the identities of those children.

2.4. Video enhancement

Image enhancement is a technique used to improve the quality of digital images by optimizing contrast, sharpness, and clarity or by emphasizing specific features within the image. The primary objective of enhancement is to optimize the visual information present in the image, making it more suitable for a specific application or analysis [23]. HE, CS, and CLAHE are included in the video-enhancing techniques in this study. By redistributing the pixel intensity, histogram equalization enhances the contrast and visibility of photographs. By analyzing the distribution of pixel values in a histogram, histogram equalization adjusts the intensity levels to create a more balanced and even distribution. This image processing method (contrast stretching) seeks to widen the gap between the intensity values of an image's darkest and lightest pixels. Meanwhile, CLAHE enhances an image's local contrast to enhance the detail in areas of an image that are too dark or too bright. Fig. 3 shows the result of contrast enhancement from the video recording frame.

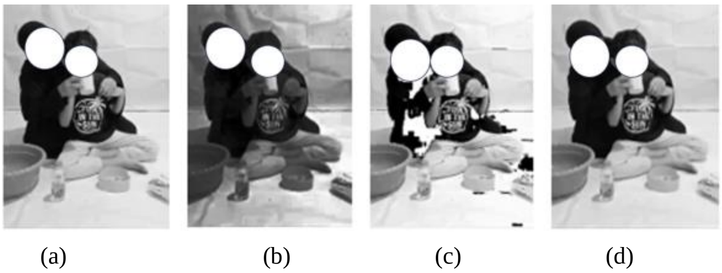


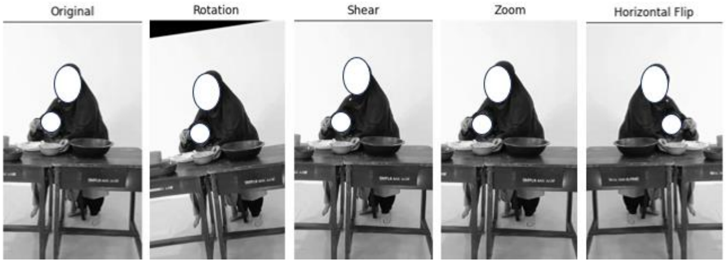
Fig. 3. Result of contrast enhancement: Grayscale (a); HE (b); CS (c); CLAHE (d)

Fig. 3. shows the Result of contrast enhancement. Fig. 3a shows the original grayscale image, with a gradient from black to white. It is the base image before contrast enhancement is applied. Fig. 3b contrast enhancement has been applied. The dark areas have become lighter, and the light areas have become darker, resulting in overall higher contrast. Fig. 3c shows further contrast enhancement has been applied, making details in the image clearer. The difference between dark and light areas is more pronounced. Fig. 3d shows the result of adaptively enhanced contrast. The details in the image are clearer, especially in areas that were previously less visible, both in the dark and light regions.

2.5. Recognizing the daily activities of children with ADS

Data splitting was performed before the recognition started. The data was split into training and testing groups at random using an 80:20 distribution ratio. The validation data was divided into 20% of the training data. Data augmentation was performed in order to expand the quantity of data that is accessible by altering images under different circumstances, such as rotation, flip, shift, and zoom. Previous studies have demonstrated that the correct augmentation techniques reduced the likelihood of overfitting and improved the system's robustness [24]. The dataset in this study was augmented using four methods: rotation, shear, zoom, and horizontal flip. The result of data augmentation is shown in Fig. 4.

The recognition of daily activities involves training a CNN model. Classification is performed using two different CNN architectures: VGG19 and MobileNetV2. VGG19 is a CNN architecture that uses very small convolutional filters (3×3) and a deep structure consisting of 19 layers [25]. The VGG19 model used fine-tuning with a pre-trained model from ImageNet, with input data dimensions of 224×224 pixels. The pre-trained feature learning layers were kept unchanged during training to prevent significant alterations to the previously learned features. In the classification layer, a flattened layer was used to modify the vector dimensions before they were input into a fully linked layer for prediction. The fully connected layer consisted of two layers, each with 4.096 neurons. The classification process concluded with predicting each image's label using a prediction layer with eight neurons for eating and six for drinking activities.



(a) (b) (c) (d) (e)

Fig. 4. Result of data augmentation: Original (a); Rotation (b); Shear (c); Zoom (d); Horizontal Flip (e)

The next phase of model training involved using MobileNetV2. MobileNetV2 is an updated version of MobileNetV1, designed as a CNN model. It is optimized for devices that have limited computational power [26]. This model leverages a pre-trained ImageNet model for feature learning layers, with input data at 224×224 pixels. In this study, the pre-trained feature learning layers were kept unchanged during training to prevent significant alterations to the previously learned features. The last convolutional layer applied Global Average Pooling (GAP) to reduce the feature dimensions from the preceding layer. The output from GAP served as the classification layer's input, consisting solely of a prediction layer with eight neurons for eating and six for drinking activities.

Table 3 lists eight proposed models, four for each of the two datasets (eating and drinking activities). The models are: (1) CLAHE with VGG19, (2) VGG19 without image enhancement, (3) CLAHE with MobileNetV2, and (4) MobileNetV2 without image enhancement. These models are applied to both the eating activities dataset and the drinking activities dataset.

Table 3. The proposed models

No of proposed model	Dataset	Image enhancement	Model
1	Eating activities	CLAHE	VGG19
2		Without	
3		CLAHE	MobileNetV2
4		Without	
5	Drinking activities	CLAHE	VGG19
6		Without	
7		CLAHE	MobileNetV2
8		Without	

2.6. Evaluation

The CNN performance is evaluated using SSIM and PSNR[27]. The former is widely used to quantify the gap between the values estimated by the model and the actual ones. In image/video processing, SSIM is used to evaluate the quality of reconstructed or compressed images/videos compared to their original versions, as indicated in the next equation [28, 29],

$$\text{SSIM} = \frac{(2\mu_x\mu_y+C_1)+(2\sigma_{xy}+C_2)}{(\mu_x^2+\mu_y^2+C_1)+(\sigma_x^2+\sigma_y^2+C_1)} \tag{1}$$

Each x and μ_x represent the original image and its mean value, respectively. Meanwhile, y and μ_y denote the modified image and its mean value. The covariance between the original and modified images is expressed as σ_{xy}. The variables C_1 and C_2 are used to stabilize the division when the denominator is weak. The variances of the original and modified images are denoted as σ_x^2 and σ_y^2, respectively.

PSNR is a metric used to assess the quality of processed images. It is measured in dB units. PSNR uses a simple pixel comparison approach, helping to evaluate coding techniques for their effectiveness. The next equation is used to calculate PSNR,

$$\text{PSNR} = 10\,.\log_{10}\frac{255^2}{\sqrt{\text{MSE}}}. \tag{2}$$

MSE indicated in the next equation. The variables m and n provide the cover image's width and height; meanwhile, *c* represents the cover image, and *s* represents the stego image:

$$\text{MSE} = \sum_{m=0}^{m}\sum_{n=0}^{n}||(c(m,n)-s(m,n)||. \tag{3}$$

The higher the PSNR value, the higher the quality, which means the signal is stronger than the noise. Likewise, the lower the value, the more distortion or loss of quality. PSNR is widely used in image and video processing, particularly in applications where fidelity is crucial, such as medical imaging, satellite imaging, and video streaming.

The True/False Positive (TP/FP) and True/False Negative (TN/FN) values represent the confusion matrix calculation. These values are used to evaluate the CNN process's performance, as shown in the confusion matrix [25]. Accuracy measures the number of correct predictions according to the actual class out of all predictions made by the model, as indicated in the equation

$$\text{Accuracy} = \frac{\text{TN + TP}}{\text{TP + FP + TN + FN}}. \tag{4}$$

Precision measures the model's performance in correctly predicting the positive class but does not include the negative class wrongly predicted as positive. A high precision value indicates that the model does not misclassify the negative class as the positive class through the calculation shown in the equation

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}}. \tag{5}$$

As shown in the Equation (6) bellow, recall gauges how well the model accurately predicts every positive class. The F1-score measures the performance of the model by calculating the precision and recall values, as shown in the Equation (7) bellow:

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}}, \tag{6}$$

$$\text{F1-score} = \frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}}. \tag{7}$$

3. Results and discussion

The methods were tested using a primary data set that included video recordings of daily activities of children with ASD conducted by therapists and parents. This testing aimed to assess the effectiveness of various contrasting methods in enhancing the quality of these video recordings and to recognize daily activities associated with ASD using CNN architecture. This architecture has successfully recognized the daily activities of children with ASD in videos by improving image quality through enhanced contrast.

3.1. Analysis of the enhanced images

The effectiveness of the investigated contrast enhancement methods is evaluated using two metrics: SSIM and PSNR. Fig. 5 presents the test results of these methods in an example video. The sample data used was a video recording named “Opening the Bottle (9).mp4”, which was extracted to produce 375 frames. Each frame was first converted to grayscale, and then three contrast enhancement techniques – HE, CS, and CLAHE – were applied to boost the contrast in each frame.

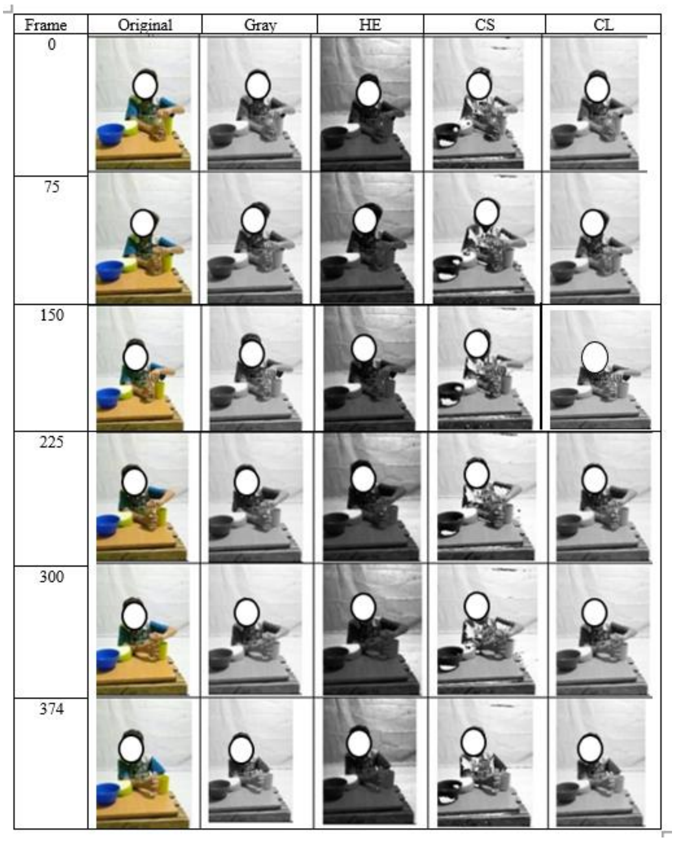


Fig. 5. Contrast enhancement results of the frames of opening the bottle (9).mp4

The metrics, SSIM and PSNR, validate the accuracy of the test results. Table 4 presents the average SSIM values and PSNR values using eating activities. Table 4 shows that the average SSIM results with HE was 0.850, CS at 0.804, and CLAHE at 0.998. The average PSNR was 17.438 utilizing HE, 11.717 CS, and 38.466 CLAHE. Table 5 displays the average SSIM and PSNR values of the drinking activities. The mean SSIM result of HE was 0.820, CS 0.803, and CLAHE 0.998. The average PSNR was 16.944 utilizing HE, 11.286 CS, and 38.296 CLAHE. The SSIM values range from –1 to 1 if 1 indicates that the two images are identical and 0 or negative values indicate very low similarity. PSNR will provide the value of the image processing outcomes in the interim. If the PSNR value is higher than 35 dB, the image is more accurate. On the other hand, poor image processing outcomes occur when the PSNR value is less than 35 dB [29]. The average PSNR above 35 dB and CLAHE are categorized as good (with an average SSIM of 0.998 and PSNR of 38.466 for eating activities and an average SSIM of 0.998 and PSNR of 38.296 for drinking activities).

Table 4. The mean values of SSIM and PSNR of the eating activities

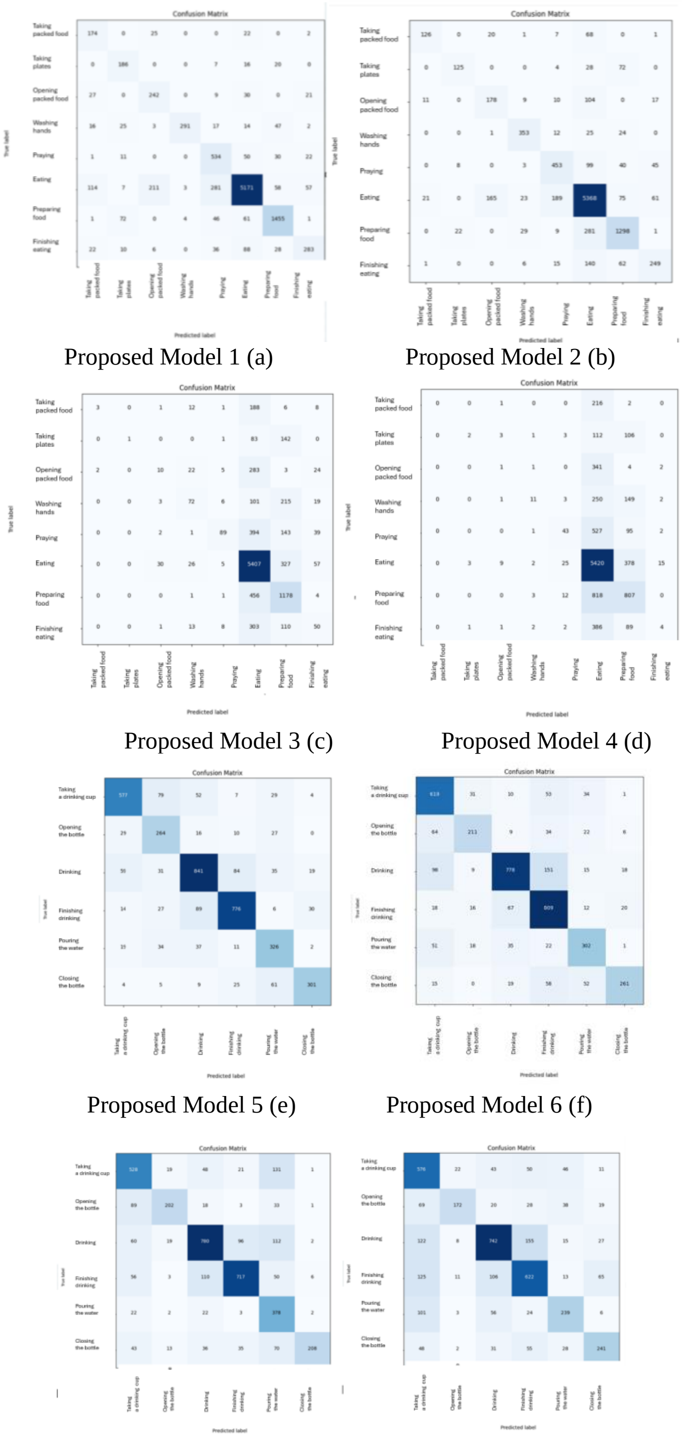
No	Activities	HE		CS		CLAHE	
		SSIM	PSNR	SSIM	PSNR	SSIM	PSNR
1	Washing hands	0.850	16.320	0.799	11.460	0.998	38.990
2	Taking a plate	0.852	16.885	0.805	11.735	0.998	39.345
3	Preparing the food	0.866	18.920	0.812	11.320	0.997	39.300
4	Taking the packed food	0.857	17.380	0.799	11.280	0.998	37.820
5	Opening the packed food	0.843	16.580	0.788	11.050	0.997	38.350
6	Praying	0.846	16.880	0.803	11.200	0.998	38.270
7	Eating	0.862	19.700	0.806	14.454	0.998	37.474
8	Finishing eating	0.821	16.840	0.817	11.240	0.998	38.180
Average		0.850	17.438	0.804	11.717	0.998	38.466

Table 5. The mean values of SSIM and PSNR of the drinking activities

No	Activities	HE		CS		CLAHE	
		SSIM	PSNR	SSIM	PSNR	SSIM	PSNR
1	Taking a drinking cup	0.828	17.100	0.809	11.220	0.998	38.105
2	Opening the bottle	0.801	16.360	0.793	11.423	0.998	38.363
3	Pouring the water	0.834	17.240	0.804	11.460	0.998	38.970
4	Drinking	0.798	17.445	0.809	11.150	0.998	38.015
5	Closing the bottle	0.840	16.575	0.795	11.260	0.998	38.285
6	Finishing drinking	0.818	16.945	0.808	11.200	0.998	38.040
	Average	0.820	16.944	0.803	11.286	0.998	38.296

3.2. Results of the recognition

This study performed image enhancement and classification experimentations for eating and drinking activities for two models (i.e., VGG19 and MobileNetV2). The classification metrics were computed to verify the suggested model’s efficacy. Fig. 6 (a)-(h) represents the confusion matrix for all the proposed models using VGG19 and MobileNetV2 models concerning the CLAHE enhancement for the videos of eating and drinking activities. We compute the additional performance measures using the Equations (3)-(6) according to the settings that the confusion matrix employed. Table 6 illustrates the performance indicators for the eating and drinking activities dataset.



Proposed Model 1 (a)

Proposed Model 2 (b)

Proposed Model 3 (c)

Proposed Model 4 (d)

Proposed Model 5 (e)

Proposed Model 6 (f)

Proposed Model 7 (g)

Proposed Model 8 (h)

Fig. 6. The confusion matrix for all proposed models

The final testing findings for each of the suggested models are shown in Table 6. The experiment’s findings show that the CLAHE and VGG19 models outperform the other models by a wide margin, with accuracy rates of 78% for the drinking datasets and 85% for the eating activities. Figs 10-13 illustrate the trade-offs between the accuracy and corresponding loss for training and validation (the proposed model using the eating activities dataset) and 14-17 (the proposed model using the drinking activities dataset), respectively.

Table 6. Performance metrics assessed for each model

No of proposed model	Performance measures			
	Accuracy	Precision	Recall	F1-score
1	0.85	0.87	0.85	0.85
2	0.83	0.83	0.83	0.82
3	0.69	0.67	0.69	0.63
4	0.64	0.55	0.64	0.55
5	0.78	0.78	0.79	0.78
6	0.76	0.76	0.77	0.76
7	0.71	0.71	0.75	0.71
8	0.66	0.66	0.67	0.66

On the other hand, MobileNetV2 with no image enhancement had the lowest accuracy, at 64 % for the eating activities dataset and 66 % for the drinking dataset. Figs 7-10 illustrate the trade-offs between training and validation accuracy as well as the corresponding losses for each (the proposed models using the eating activities dataset) and Figs 11-14 (the proposed models using the drinking activities dataset), respectively.

This study’s findings demonstrate that VGG19 offers more promising results in recognizing the daily activities of children with ASD, considering the image enhancement (CLAHE) for the eating activities dataset. In comparison, the performance of the CLAHE and VGG19 using the drinking activities dataset is the second best. The MobileNetV2 without image enhancement results in the lowest score.

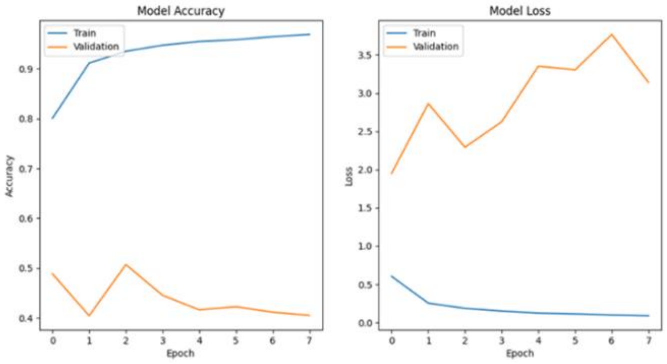


Fig. 7. The comparison between training and validation for CLAHE and VGG19 using eating activities: accuracy (a); loss (b)

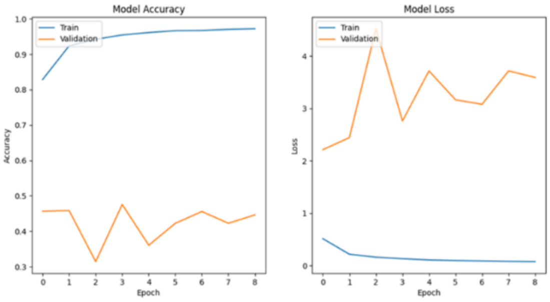


Fig. 8. The comparison between training and validation for VGG19 using eating activities: accuracy (a); loss (b)

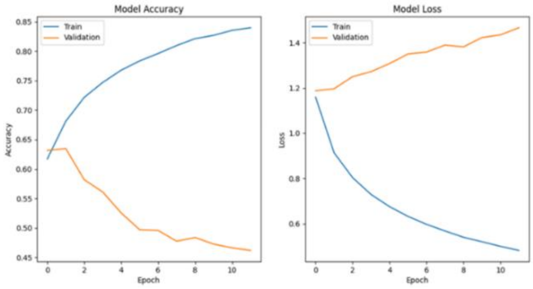


Fig. 9. The comparison between training and validation for CLAHE and MobileNetV2 using eating activities: accuracy (a); loss (b)

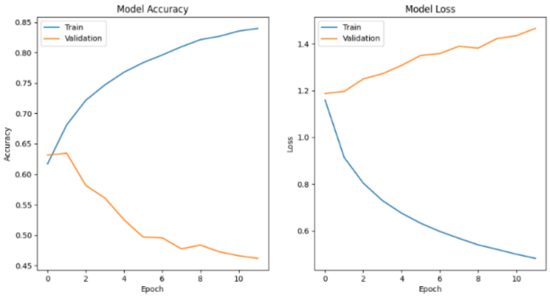


Fig. 10. The comparison between training and validation for MobileNetV2 using eating activities: accuracy (a); loss (b)

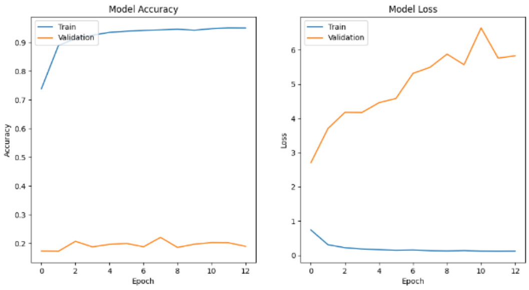


Fig. 11. The comparison between training and validation for CLAHE and VGG19 using drinking activities: accuracy (a); loss (b)

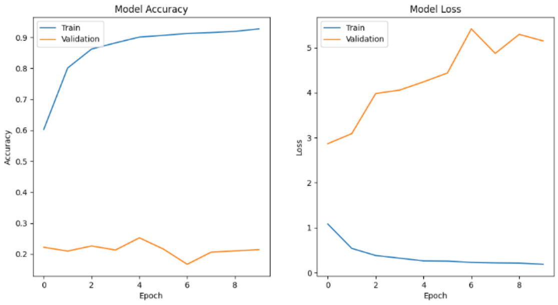


Fig. 12. The comparison between training and validation for VGG19 using drinking activities: accuracy (a); loss (b)

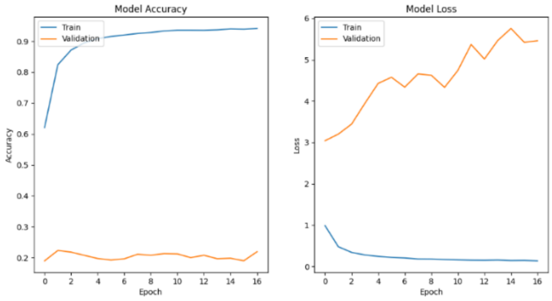


Fig. 13. The comparison between training and validation for CLAHE and MobileNetV2 using drinking activities: accuracy (a); loss (b)

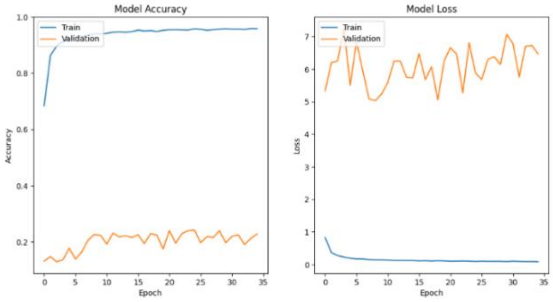


Fig. 14. The comparison between training and validation for MobileNetV2 using drinking activities: accuracy (a); loss (b)

3.3. Comparison of proposed methods

The usefulness of our conclusions from the proposed model is then verified by comparing the results with those of other state-of-the-art methods. Previous studies [16, 17, 30] discussed the diagnosis of ASD using secondary data obtained from public websites. In those studies, the diagnosis was carried out using a Support Vector Machine (SVM) and CNN architecture. However, these studies did not include image processing techniques, such as image enhancement, to diagnose ASD.

The main contribution of this study is the use of primary data in the form of videos of daily activities, such as eating and drinking, recorded directly by therapists or parents. In addition, this study contributes to selecting the optimal image enhancement strategy for various datasets to improve the recognition of children with ASD's performance in completing daily activities. The third contribution is the recognition of children with ASD's daily activities using the CNN architecture, which has demonstrated success across various datasets with enhanced video quality (improved contrast).

Table 7 compares the performance of the techniques based on accuracy and baseline architecture. Table 7 demonstrates that image enhancement improves the CNN model's performance significantly. The accuracy of CLAHE and VGG19 far surpasses that of VGG19 without enhancement. Similarly, CLAHE and MobileNetV2 show significantly higher accuracy than MobileNetV2 without enhancement. Therefore, these results show that combining transfer learning with image enhancement substantially boosts the accuracy of recognizing the daily activities of children with ASD.

Table 7. Comparison of proposed methods

Reference	Dataset	Baseline architecture	Accuracy	Year
[30]	functional Magnetic Resonance Imaging (fMRI)	SVM	70%	2019
[16]		CNN	70.22%	2019
[17]		CNN	78%	2020
Proposed Model 1	Eating activities (Primary data)	CLAHE + VGG19	85%	2024
Proposed Model 2		VGG 19	83%	2024
Proposed Model 3		CLAHE + MobileNetV2	69%	2024
Proposed Model 4		MobileNetV2	64%	2024
Proposed Model 5	Drinking activities (Primary data)	CLAHE + VGG19	78%	2024
Proposed Model 6		VGG19	76%	2024
Proposed Model 7		CLAHE + MobileNetV2	71%	2024
Proposed Model 8		MobileNetV2	66%	2024

3.4. Discussion

Tables 4 and 5 demonstrate that CLAHE is an effective image enhancement technique, as the SSIM values range from –1 to 1 if 1 indicates that the two images are identical and 0 or negative values indicate very low similarity. Specifically, the SSIM is 0.998 for eating and drinking. Additionally, the PSNR values are above 35, with 38.47 for eating and 38.40 for drinking. In contrast, a higher PSNR value above 35 dB suggests more accurate image processing, while a value below 35 dB indicates poorer outcomes [22]. Based on these results, CLAHE is selected as an image enhancement technique to recognize the daily activities of children with ASD.

Table 7 demonstrates that the CNN model utilizing image enhancement achieves higher accuracy. The CLAHE and VGG19 model provides results with an accuracy of 85% and VGG without image enhancement of 83 % for the eating activities dataset, while the drinking activities of 78 % for CLAHE and VGG19 model and 76% for VGG19 without image enhancement. This improvement is attributed to image enhancement techniques such as CLAHE, which enhance the contrast and clarity of image features, making it easier for CNN to detect and learn important features, leading to better classification performance.

Table 7 reveals that VGG19 achieves slightly better accuracy than MobileNetV2, though the difference is insignificant. The CLAHE and VGG19 model yields an accuracy of 83% for eating activities, whereas CLAHE and MobileNetV2 achieve 71% for drinking activities. VGG19's greater depth, with its 19 layers, enables it to capture more complex features and patterns, resulting in superior performance for tasks demanding high accuracy [25].

HAR is often carried out through the analysis of videos or images. By utilizing image enhancement techniques, the quality of images from video calls, recordings sent by parents, or surveillance videos in schools can be improved in terms of contrast. This facilitates algorithms in detecting certain movements or activities, such as hand gestures, facial expressions, or body posture. This is particularly crucial in ASD research, where children's behavior and movements can be studied to recognize daily activities related to ASD.

In Indonesia, access to diagnostic and therapeutic services is often limited, especially in remote areas. By using image enhancement, the quality of images from video calls or recordings sent by parents can be enhanced, allowing experts in major cities to conduct remote analyses more accurately. Clearer and more informative images can help the public understand the importance of early detection and support for children with ASD, ultimately reducing negative stigma.

4. Conclusion

This study presents a method for determining the optimal image enhancement strategy for various datasets using CNN to improve the recognition of the daily activities of children with ASD. CLAHE is an effective image enhancement technique in ASD's daily activities recognition, achieving an SSIM of 0.998 and a PSNR of 38.466 for the eating activities dataset and an SSIM of 0.998 and a PSNR of 38.296 for the drinking activities dataset**.**

The integration of CNN architectures with CLAHE-enhanced video data, utilizing modified VGG19 and MobileNetV2 models with customized dense networks and classification layers, demonstrated promising results. The VGG19 model with CLAHE enhancement achieved classification accuracies of 85% for eating activities and 78% for drinking activities. Future research will benefit from seeking to create a new real-time model capable of recognizing ASD's daily behaviors.

Acknowledgments**:** The authors would like to thank the Universitas Airlangga who had trusted and funded the research entitled Recognition of Daily Activities in Children with Autism Spectrum Disorder (ASD) Using Convolutional Neural Networks (CNN) with contract number: 1663/UN3.FST/PT.01.03/2024. We express our gratitude for facilitating this research so that it can be completed without any significant difficulties.

References

1. Kamp-Becker, I. Autism Spectrum Disorder in ICD-11 – A Critical Reflection of its Possible Impact on Clinical Practice and Research. – Molecular Psychiatry, Vol. **29**, 2024, pp. 633-6387.

2. Sulkes, B., S. Definition of Developmental Disorders. **https://www.msdmanuals.com/home/children-s-health-issues/learning-and-developmental-disorders/definition-of-developmental-disorders**
3. Daulay, N. Parenting Stress of Mothers in Children with Autism Spectrum Disorder: A Review of the Culture in Indonesia. – In: Proc. of International Conference on Southeast Asia Studies, 2018.
4. Henry, M., E. Living Life Like It's Golden with Disability: Case Studies of Independent Living. 2018.
5. Mohd Kamil, N. K., A. S. Amin, N. Md Akhir, A. R. Ahmad Badayai, I. Mohd Zambri, R. Sutan, K. F. Khairuddin, W. A. Wan Abdullah. Independent Living Skills Needed by Students with Special Educational Needs (SEN) Towards Inclusive Education: A Systematic Literature Review. – Specialists Ugdym, Vol. **1**, 2023, No 44, pp. 610-623.
6. Volkmar, F. R. Encyclopaedia of Autism Spectrum Disorders. Springer, Switzerland, 2021.
7. Pramardika, D., D. Susanti, E. Fitriana. Analisis Pola Makan Anak Autis Yayasan Tongkat Musa Indonesia ABK Bangun Rejo Kabupaten Kutai Kartanegara Tahun 2019. – Bunda Edu-Midwifery Journal, Vol. **2**, 2019, No 1, pp. 18-24.
8. Kurniati, L. Modul Guru Pembelajar SLB Autis. PPPPTK TK DAN PLB, Bandung, 2016.
9. Beddiar, D. R., B. Nini, M. Sabokrou, A. Hadid. Vision-Based Human Activity Recognition: A Survey. – Multimed Tools Appl Journal, Vol. **79**, 2020, pp. 30509-30555.
10. Su, X., H. Tong, P. Ji. Activity Recognition with Smartphone Sensors. – Tsinghua Science and Technology, Vol. **19**, 2014, No 3, pp. 235-249.
11. Jain, A., V. Kanhangad. Human Activity Classification in Smartphones Using Accelerometer and Gyroscope Sensors. – IEEE Sensors Journal, Vol. **18**, 2018, No 3, pp. 1169-1177.
12. Yao, G., T. Lei, J. Zhong. A Review of Convolutional-Neural-Network-Based Action Recognition. – Pattern Recognition Letters, Vol. **118**, 2019, pp. 14-22.
13. Dhillon, A., G. K. Verma. Convolutional Neural Network: A Review of Models, Methodologies and Applications to Object Detection. – Progress in Artificial Intelligence, Vol. **9**, 2020, No 2, pp. 85-112.
14. Alzubaidi, L., J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaria, M. A. Fadhel, M. Al-Amidie, L. Farhan. Review of Deep Learning: Concepts, CNN Architectures, Challenges, Applications, Future Directions. – Journal of Big Data, Vol. **8**, 2021, pp. 1-74.
15. Bhatt, D., C. Patel, H. Talsania, J. Patel, R. Vaghela, S. Pandya, K. Modi, H. Ghayvat. CNN Variants for Computer Vision: History, Architecture, Application, Challenges and Future Scope. – Electronics, Vol. **10**, 2021, No 2470, pp. 1-28.
16. Haweel, R., A. Shalaby, A. Mahmoud, N. Seada, S. Ghoneims, M. Ghazal, M. F. Casanova, G. N. Barnes, A. El-Baz. A Robust DWT – CNN-Based CAD System for Early Diagnosis of Autism Using Task-Based fMRI. – Medical Physics, Vol. **48**, 2020, No 5, pp. 2315-2326.
17. Sherkatghanad, Z., M. Akhondzadeh, S. Salari, M. Zomorodi-Moghadam, M. Abdar, U. R. Acharya, R. Khosrowabadi, V. Salari. Automated Detection of Autism Spectrum Disorder Using a Convolutional Neural Network. – Front Neurosci, Vol. **13**, 2020, pp. 1-17.
18. Mitschke, N., Y. Ji, M. Heizmann. Task Specific Image Enhancement for Improving the Accuracy of CNNs. – In: Proc. of 10th International Conference on Pattern Recognition Applications and Methods, 2021, pp. 174-181.
19. Ferdinand, V., A. Henry, G. E. Nawir, V. Anderies., A. Gunawan. Effect of Image Enhancement in CNN-Based Medical Image Classification: A Systematic Literature Review. – In: Proc. of 5th International Conference on Information and Communications Technology, 2022, pp. 87-92.
20. Gonzalez, R. C., R. E. Woods. Digital Image Processing. New York, Pearson, 2018.
21. Werdiningsih, I., I. Puspitasari, R. Hendradi. Analysis and Techniques of Enhancing the Video Quality of Children with Autism Spectrum Disorder's Daily Activities – In: Proc. of 24th International Seminar on Intelligent Technology and Its Applications (ISITIA'24), 2024, pp. 621-626.

22. Mustaghfirin, F., H. Erwin, K. Putra, U. Yanti, R. Ricadonna. The Comparison of Iris Detection Using Histogram Equalization and Adaptive Histogram Equalization Methods. – In: Proc. of International Conference on Information System Computer Science and Engineering, 2019.

23. Qi, Y., Z. Yang, W. Sun, M. Lou, J. Lian, W. Zhao, X. Deng, Y. Ma. A Comprehensive Overview of Image Enhancement Techniques. – Archives of Computational Methods in Engineering, Vol. **29**, 2022, pp. 583-607.

24. Lu, P., B. Song, L. Xu. Human Face Recognition Based on Convolutional Neural Network and Augmented Dataset. – Systems Science & Control Engineering, Vol. **9**, 2021, No 2, pp. 29-37.

25. Rao Killi, C. B., N. Balakrishnan, C. S. Rao. Deep Fake Image Classification Using VGG-19 Model. – International Information and Engineering Technology Association, Vol. **28**, 2023, No 2, pp. 509-515.

26. Rusia, M. K., D. K. Singh. A Color-Texture-Based Deep Neural Network Technique to Detect Face Spoofing Attacks. – Cybernetics and Information Technologies, Vol. **22**, 2022, No 3, pp. 127-145.

27. Habiban, M., F. R. Hamade, N. A. Mohsin. Hybrid Edge Detection Methods in Image Steganography for High Embedding Capacity. – Cybernetics and Information Technologies, Vol. **24**, 2024, No 1, pp. 157-170.

28. Sara, U., M. Akter, M. S. Uddin. Image Quality Assessment through FSIM, SSIM, MSE and PSNR – A Comparative Study. – Journal of Computer and Communications, Vol. **7**, 2019, No 3, pp. 8-18.

29. Erwin, D. R. Ningsih, Improving Retinal Image Quality Using the Contrast Stretching, Histogram Equalization, and CLAHE Methods with Median Filter. – International Journal of Image, Graphics and Signal Processing, Vol. **12**, 2020, No 2, pp. 30-41.

30. Huan, B., K. Zhang, R. Sanchez-Romero, J. Ramsey, M. Glymoury, C. Glymoury. Diagnosis of Autism Spectrum Disorder by Causal Influence Strength Learned from Resting-State fMRI Data. – Imaging and Signal Analysis Journal, Vol. **1**, 2019, pp. 237-267.

*Received: 11.09.2024, First Revision: 04.12.2024, Second Revision: 15.12.2024,
Accepted: 20.12.2024*