

# Classifying the suitability of educational videos for attention deficit hyperactivity disorder students with deep neural networks

Alshefaa Emam<sup>1</sup>, Eman Karam Elsyed<sup>2,3</sup>, Mai Kamel Galab<sup>1</sup>

<sup>1</sup>Department of Computer Science, Faculty of Computers and Artificial Intelligence, Benha University, Benha, Egypt

<sup>2</sup>Department of Mathematics, Faculty of Science, Al-Azhar University (Girls), Cairo, Egypt

<sup>3</sup>Department of Computer Science, College of Information Technology, Misr University for Science and Technology (MUST), Giza, Egypt

## Article Info

### Article history:

Received Jan 29, 2025

Revised Jun 24, 2025

Accepted Jul 3, 2025

### Keywords:

ADHD online instructor dataset

Attention deficit hyperactivity disorder

Convolutional neural network

Deep learning

Video classification

## ABSTRACT

This paper presents a comprehensive deep learning-based system to evaluate the educational videos' suitability for students with attention deficit hyperactivity disorder (ADHD). Current methods frequently ignore important instructional elements that are necessary for improving learning experiences for students with ADHD, such as instructor hand movements, video length, object variety, and audio-visual quality. We emphasize two key issues for how to address these difficulties, first, we present the ADHD online instructor (AOI) dataset, a particular benchmark for assessing instructional hand movement in video suitability to solve the absence of a reference dataset for classifying educational videos relevant to ADHD. Second, the system includes creating an enhanced multitask deep learning model that increases classification accuracy by using task-specific enhancements and optimized architectures. This solves the requirement for a strong model that can distinguish between suitable and unsuitable instructional content. Comprehensive tests using pretrained convolutional neural network (CNN) models indicate that the enhanced VGG16 model outperforms baseline methods by achieving a highest accuracy of 97.84%. The results highlight the value of integrating deep learning methods with structured benchmark datasets, exposing up the path to more resilient and flexible instructional materials designed for students with ADHD.

*This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.*



## Corresponding Author:

Alshefaa Emam

Department of Computer Science, Faculty of Computers and Artificial Intelligence, Benha University  
Benha, Egypt

Email: shefaaemam97@gmail.com

## 1. INTRODUCTION

With online learning growing at a rapid pace, the United Nations 2030 Agenda emphasizes the urgent need for high-quality, inclusive, and accessible education, including students with attention deficit hyperactivity disorder (ADHD). However, these students frequently find internet educational videos to be quite difficult, mostly because of the length, complexity, and general quality of the videos. Students with ADHD have special learning needs that are often overlooked by traditional teaching techniques and digital tools, which lowers their efficacy and engagement [1]. The effects of many components of instructional films for students with ADHD, such as item diversity, video length, audio and visual quality, and the teacher hand movement, have been the subject of extensive research. Hand movement are among the most important nonverbal clues that can improve comprehension for students with ADHD. Research shows that using

gestures in the classroom increases the likelihood that students with ADHD will learn and participate [2]. This suggests that watching movies with these cues may help students learn more effectively and pay attention for longer periods of time. Additionally, the length of movies is important for maintaining youngsters with ADHD's attention. Videos longer than fifteen minutes have been shown to reduce viewer attention and memory recall [3]. Thus, setting sensible time restrictions for videos is crucial to maximizing the educational process. Additionally, improving attention and engagement requires high-quality audio and visual content. Poor quality content can hinder understanding and be distracting, especially for children who have trouble focusing [4]. The contributions of the paper can be summarized in the following point: i) A new system of evaluating educational videos that are suitable for students with ADHD is presented; ii) The ADHD online instructor (AOI) dataset was created, regarding the absence of a dataset aimed at this purpose, it includes 10,600 photos that highlight important elements such as the instructor's hand movements; iii) A five-phases content the suitability evaluation system was created, to make sure that it meets the needs of students with ADHD, this system efficiently analyses and categorizes video content using the enhanced VGG16 model in combination with the YOLO and OpenCV libraries.

The paper is organized as follows: In section 2, examines the work relating to this paper. Section 3 describes the proposed ADHD system. In section 4, the results of the proposed system are examined. Section 5 provided a future direction for the suggested system and summarized the findings of this work.

## 2. RELATED WORK

The literature on deep learning-based video classification for ADHD students is reviewed in this section. It addresses five main issues: attempts to categories videos according to their suitability and quality; difficulties ADHD students face when learning online; the application of deep learning to video analysis; methods such as feature extraction and data augmentation; and comparisons with current approaches. These fields show the gaps in our existing understanding and give our research context. The impact of visual and auditory components on learning outcomes is highlighted by studies examining the function of sensory processing in educational settings. For example, studies show that movies with a moderate amount of visual complexity might increase learners' engagement by giving them enough stimulation without being too demanding [5]. But there is not a thorough examination of ADHD-specific issues like attention management in this study. Similarly talks on how sensory processing problems affect learning and concentration more generally, but it does not offer any practical solutions for ADHD children [6].

Additional studies examine how people with ADHD struggle to integrate audiovisual information, highlighting how problems with sensory integration can impair their interest and learning [5]. This is furthered by study [7], who investigate the difficulties faced by kids with ADHD who wear bilateral hearing aids, highlighting the necessity of specialized teaching strategies. Despite these realizations, research on sensory processing has not been widely incorporated into useful instructional frameworks. Research has also been done on the benefits of immersive technology, such augmented reality (AR), in video-based learning. By offering immersive and engaging experiences, AR can improve understanding and engagement for kids with ADHD, as shown by study [8]. However, this study mostly concentrates on the advantages of AR without discussing the difficulties of incorporating these technologies into current educational institutions or their scalability for wider implementation. Although these studies highlight variables that support students with ADHD, they frequently focus on these components separately, ignoring a comprehensive approach.

Numerous studies have examined gestures as a teaching tool, emphasizing how crucial they are for improving understanding and communication in classroom environments. Suggest a deep learning-based hand gesture detection system that enhances teacher-student interactions by accurately classifying gestures in real-time classroom settings [2]. However, the study is limited by its focus on technological feasibility without evaluating its impact on long-term learning outcomes. FREITAS and investigate several forms of gestures and their roles in classroom training, stressing their usefulness in multimodal teaching [9]. Yet, this study overlooks the limitations of standardizing gesture use across varied teaching styles. More comprehensive review emphasizes the benefits of gestures for engagement, memory retention, and comprehension while urging more study on how best to employ them in instructional tactics [10]. In order to improve 3D spatial cognition, provide "Air Anatomy," a revolutionary method of teaching spatial anatomy in medical education through hand movements [11]. Despite being novel, this approach is specialized and might not transfer well to other fields. With an emphasis on linguistic development, examine how deliberate gestures improve speaking and listening abilities [12]. Nevertheless, there is no discussion of the study's relevance to non-linguistic subjects or situations including ADHD. The difficulties that students with ADHD encounter in conventional classroom settings highlight the necessity of focused interventions. Teachers ignorance of ADHD has an impact on classroom management and instructional efficacy [13]. They advocate for increased awareness and training to improve teaching methods. However, Multimedia content that is

tailored to the learning needs of students with ADHD is crucial because these students frequently struggle with standard teaching techniques. Visual materials have the potential to improve comprehension and attention, but most study only looks at their individual components, which limits our knowledge of how successful they work together. Furthermore, the existing research does not provide useful frameworks or recommendations to assist instructors in choosing relevant resources for students with ADHD. This paper fills these gaps by putting forth an AI-based classification system that assesses educational videos according to how suitable they are for students with ADHD. The suggested system provides objective and consistent evaluations of ADHD-friendly features by automating the evaluation process using deep learning techniques and incorporating knowledge from AR-based teaching. This system gives teachers access to excellent, customized video content, which eventually enhances the educational experiences of students with ADHD. Through systematic and verified assessment, the ADHD video classifier aims to function as a comprehensive solution that serves the various needs of these students.

### 3. PROPOSED ADHD SYSTEM

The two main phases of the suggested ADHD video classifiers are data preprocessing, training, and decision, as shown on Figure 1. The subsections that follow provide specifics on each phase. When determining if educational videos are suitable for students with ADHD, several elements need to be carefully considered during the pre-processing phase. As shorter videos are better at holding the attention of students with ADHD, the procedure starts with uploading the video and evaluating its length. Videos that go above a specific amount of time are not included. After the length of the video has been determined, its quality is evaluated by looking at both the audio and the visuals to make sure there is no background noise and clarity [14], [15]. If the video satisfies these standards for quality, a sample segment is taken out for additional examination. The video's frames from various sections are included here, and their content richness and engagement are assessed. To assess the video's complexity and make sure it has a diversity of items, object detection algorithms are performed on the crucial frames [16]. The video is considered inappropriate if any frame fails this test. The frequency and level of hand movement being more than its stillness in the video are the final factors considered, with videos with insufficient motions being disqualified. A video cannot be deemed suitable for ADHD students unless all these tests are completed. Only the most appropriate and effective content is chosen thanks to this thorough screening process.

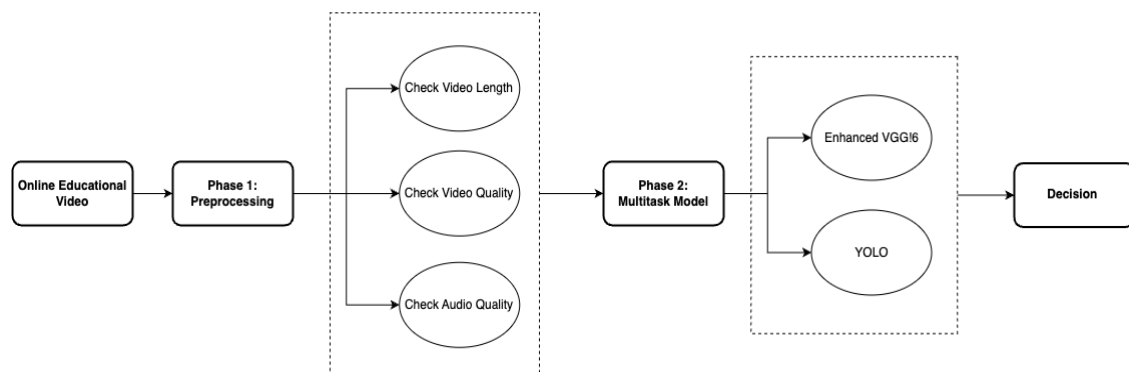


Figure 1. Workflow for ADHD video classification using multi-task models

#### 3.1. Dataset description

The suitability of online instructional videos for students with ADHD is categorized using the AOI dataset [17], which is the subject of this research. 10,600 photos make up this dataset; 5,300 of them are from videos that were judged appropriate for students with ADHD, and the remaining 5,300 are from videos that were ruled inappropriate. The process of creating the dataset started with gathering 100 YouTube videos, of which 50 were deemed appropriate for ADHD students and the other 50 inappropriate. Frames were retrieved from these videos; 14,321 frames from appropriate videos and 28,241 frames from inappropriate videos were obtained. 10,600 photos with an emphasis on frames in which the instructor's hand was moving were carefully chosen to create a balanced dataset. By capturing the crucial elements that impact understanding and engagement for ADHD students, this selection system guarantees a balanced dataset. With the help of the AOI dataset, a system that can be appropriate for students with ADHD as well as non-ADHD students can be created, improving the efficacy and quality of online learning materials for a range of student populations.

### 3.2. Data preprocessing phase

The hundred instructional videos in the ADHD video classifier dataset were hand-picked from YouTube and are categorized as "suitable" or "unsuitable" for students with ADHD. All videos were converted to frames; 14,321 frames from appropriate videos and 28,241 frames from inappropriate videos were obtained. To guarantee that pixel values are scaled to a constant range, typically between 0 and 1, all images in the training, validation, and test sets are normalized. To further differentiate the sorts of content in each category, Figure 2, the first Figure 2(a) displays images from appropriate videos, while the second Figure 2(b) plays images from inappropriate videos.

This normalization aids in accelerating the convergence of model training. The training, validation, and test sets of the dataset were split apart. Seventy percent of the photos from both categories are in the training set, fifteen percent are in the validation set, and the remaining fifteen percent are in the test set. This distribution guarantees a systematic approach to model evaluation and training. To guarantee balanced representation, 5,300 frames from each category were selected out of a total of 10,600 frames in the dataset for model training. For every category, the dataset was divided into three subsets: 795 images for testing, 9,010 images for training, and 795 photos for validation, as shown in Figure 3.

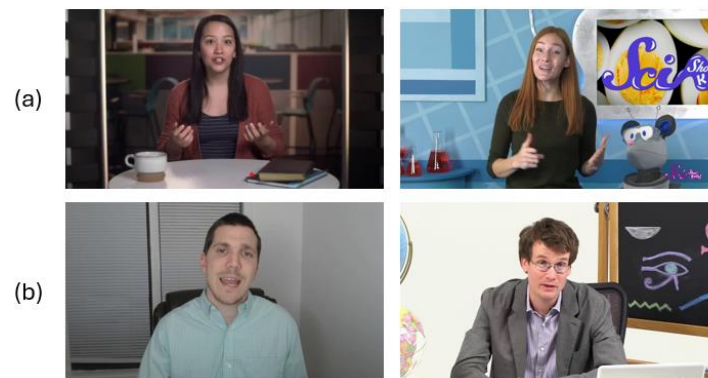


Figure 2. Few samples of (a) suitable, and (b) unsuitable in the ADHD dataset

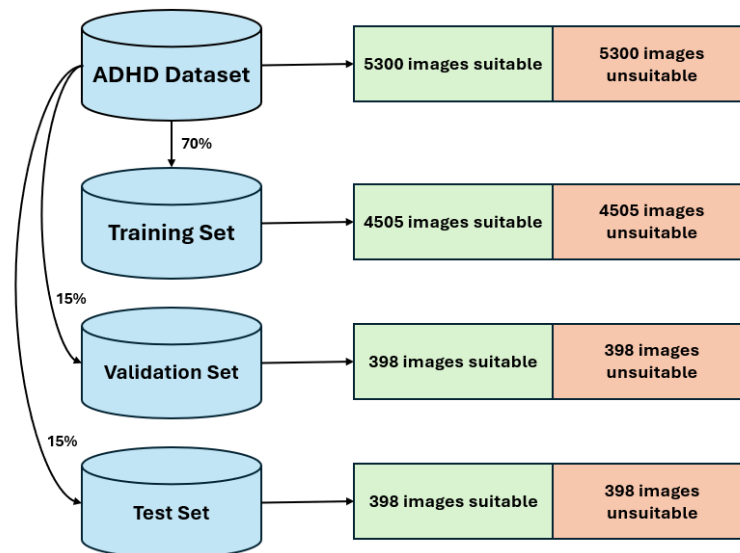


Figure 3. Illustrative structures of the data distribution in the training set, validation set, and test set

### 3.3. Training phase

Video content has been evaluated for ADHD the suitability using a variety of methods and resources. MoviePy and PyDub assessed audio quality, while OpenCV managed video length, visual quality, frame interval computations, and key frame extraction. Wget downloaded the YOLO COCO names file,

Shutil compressed files, and the OS library handled directories. For object detection, frame analysis, and loading the YOLO model, OpenCV and NumPy were applied. The analysis process was expedited by these technologies. It describes the exact software used at each phase, such as object detection, frame extraction, data compression, visual and audio quality tests, and video length verification.

This ensures a thorough and effective analysis of the video material. The VGG16 architecture was utilized as the foundation network for the training phase of the ADHD video classification model due to its established performance in image classification tasks. This architecture was pre-trained on the ImageNet dataset. The initial freezing of the convolutional layers of VGG16 allowed the training process to concentrate solely on the newly added fully connected (FC) layers, which were specifically designed for categorizing video content related to ADHD [18]. This strategy allowed the new layers to adapt to the characteristics of the ADHD dataset by utilizing the powerful features identified from the ImageNet dataset. To perform fine-tuning, the last four layers of the convolutional basis were then unfrozen [19].

The model performed better by modifying the more complex, task-specific attributes that are essential for the ADHD classification task than it did with the general features it had gained from ImageNet. The training process of the model used ten epochs, 64-batch size, and 1e-4 learning rate. Advanced training techniques, such as model checkpoint and early stopping, were applied to improve performance. The model version with the highest validation accuracy was saved by model checkpoint to ensure that the best-performing model was kept for testing. Early Stopping, which terminates training as soon as validation accuracy ceases to improve, promoted the model's generalization to new data and prevented overtraining on noise. Thus, overfitting was avoided. The VGG16-based ADHD classifier was trained using all the hyper-parameters listed in Table 1.

The original VGG16 model and the enhanced version created especially for categorizing educational videos for students with ADHD are contrasted in this section. The changes were made to the model in order to adjust it for binary classification (suitable vs. unsuitable) according to the particulars of the dataset. The main variations in model architecture, training techniques, and the performance gains attained following these adjustments are highlighted in the Table 2.

Table 1. Values assigned to the hyper-parameters of the adapted VGG16 model

Hyper-parameter	Corresponding setting
Input size	150
Count of units in the first FC layer	60
The first FC layer's activation function	ReLU
Count of units in the second FC layer	2
The second FC layer's activation function	SoftMax
Loss function	Categorical cross-entropy
Batch size (BS)	64
Optimization function	Adam
Initial learning rate (LR)	1e-4

Table 2. Comparison of original and modified VGG16 models for ADHD video classification

Aspect	Original VGG16	Modified VGG16	Reason for modification
Purpose	General image classification (1,000 categories).	Binary classification of ADHD video suitability.	Task-specific focus for ADHD dataset.
Classifier layers	Three fully connected layers.	Two layers (60 units + 2 units SoftMax).	Simplified for binary classification.
Flatten layer	Present after convolutional layers.	Added after convolutional layers.	Compatibility with modified classifier.
Layer freezing	None	Convolutional layers frozen.	Retains ImageNet features, reduces overfitting.
Fine-tuning	Not applied	Fine-tuned last four convolutional layers.	Adapts to ADHD-specific data.
Optimization	SGD with variable learning rate.	Adam with learning rate of 1e-4.	Better suited for smaller datasets.
Overfitting prevention	None	Early stopping and model checkpoint.	Reduces overfitting with limited data.
Performance	Accuracy: 87%.	Accuracy: 97.93%.	Enhanced performance for ADHD classification task.

#### 4. EXPERIMENTAL RESULTS AND DISCUSSION

The results and learnings from the assessment of the ADHD video classifier are shown in this section. The data preprocessing and training phases of the analysis are covered in a way that emphasizes the procedures and efficacy of the system. The objective is to evaluate the system's efficacy in identifying instructional videos that are appropriate or inappropriate for students with ADHD and to present a summary of the methodologies and findings.

*Classifying the suitability of educational videos for ... (Alshefaa Emam)*

#### 4.1. Results of training phase

The model's performance on a test set was evaluated using a few significant indicators, which are summarized in Table 3. The evaluation findings demonstrated the model's exceptional memory, accuracy, balanced accuracy, and confusion matrix, which demonstrated how well it classified video content related to ADHD [20]. The specificity score shed more light on how well the program could identify unfavorable occurrences [21], [22]. Accuracy is defined as (1): TP stands for true positives when actual and predicted are suitable, TN for true negatives when actual and predicted are unsuitable, FP for false positives when actual and predicted are unsuitable, and FN for false negatives when actual and predicted are unsuitable:

$$\text{Accuracy (\%)} = \left( \frac{TP + TN}{TP + TN + FP + FN} \right) \quad (2)$$

$$\text{Precision (\%)} = \left( \frac{TP}{TP + FP} \right) \quad (3)$$

$$\text{Recall (\%)} = \left( \frac{TP}{TP + FN} \right) \quad (4)$$

$$\text{F1-Score (\%)} = 2 \times \left( \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \right) \quad (5)$$

$$\text{Balanced Accuracy (\%)} = \left( \frac{\text{Sensitivity} + \text{Specificity}}{2} \right) \quad (6)$$

$$\text{Specificity (\%)} = \left( \frac{TN}{TN + FP} \right) \quad (7)$$

The confusion matrix, shown in Figure 4, provides a visual representation of the model's performance by displaying the distribution of true positives, true negatives, false positives, and false negatives [23]. The model's capacity to distinguish between suitable and inappropriate video content for ADHD is shown in this matrix. The training and evaluation results show how well the modified VGG16 model can identify information about ADHD in videos. By carefully combining the freezing and fine-tuning of the VGG16 layers with the application of state-of-the-art techniques such as model checkpoint and early stopping, a very accurate and reliable model has been generated. Furthermore, the graph of accuracy and loss for the training and test sets over the epochs is displayed in Figure 5 [24].

Table 3. Performance results of the ADHD classifier on the test set

Performance metric	Value
Accuracy	97.93%
Average Sensitivity (Recall)	97.11%
Average Precision	98.18%
Average F1-Score	97.64%
Average Specificity	98.57%
Balanced Accuracy	97.84%

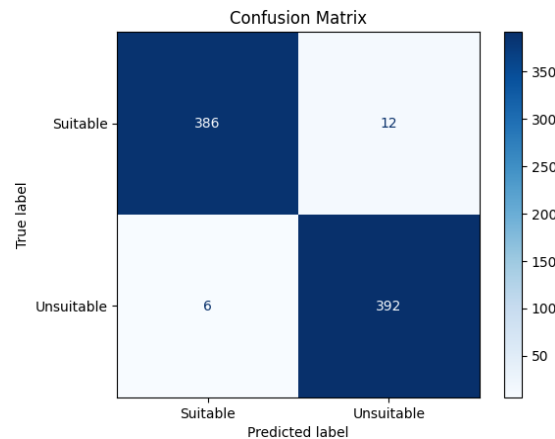


Figure 4. Confusion matrix of the ADHD video classifier

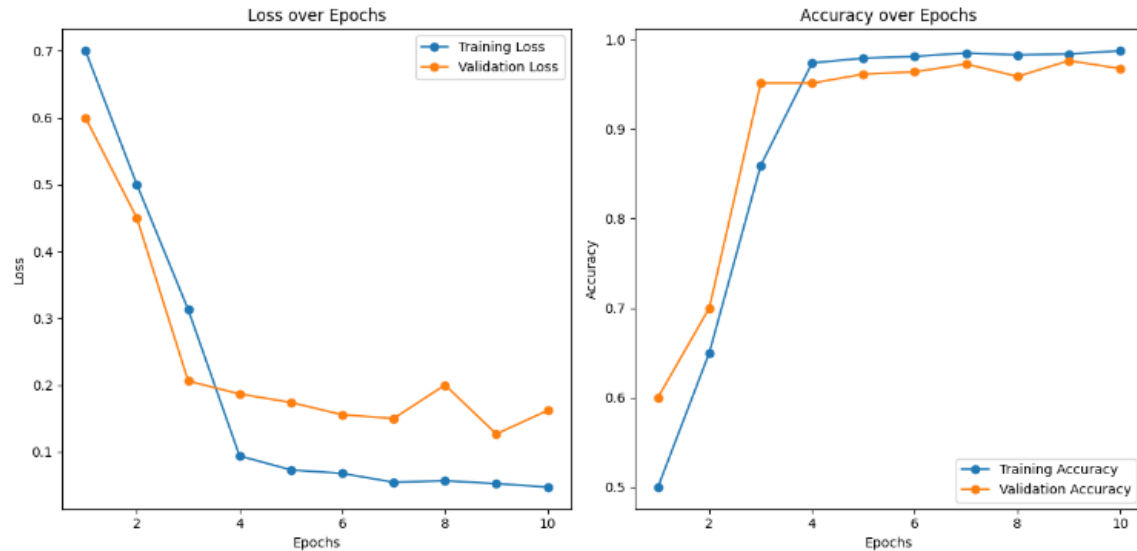


Figure 5. Training and testing loss and accuracy over epoch

#### 4.2. Results comparison with existing methods phase

Apart from the video evaluation, we also evaluated the efficacy of four well-known deep learning models on our dataset: VGG16, ResNet50, InceptionV3, and EfficientNetB0. The results are compiled in Table 4 [25]–[29]. VGG16 was chosen because of its proven ability to classify images with high accuracy 97.93% in training and 97.73% in validation. ResNet50 achieved a validation accuracy of 78.67%, but it trailed VGG16 by a small margin. ResNet50 is renowned for its creative usage of skip connections to alleviate the vanishing gradient problem.

Table 4. Comparison results with other of deep learning models

Model	Training accuracy (%)	Training loss	Validation accuracy (%)	Validation loss
Enhanced VGG16	97.93	0.0474	97.73	0.1620
ResNet50	75.52	0.5486	78.67	0.4871
InceptionV3	75.52	0.5486	78.67	0.4871
EfficientNetB0	48.99	1.0541	50.00	0.6932

Table 5 shows a comparison of two recent research with the proposed model. Importantly, the analysis of hand movement as a visual signal for categorization tasks is a shared basis among all three studies. In contrast to the first study [29], which used hand movement in radiographic imaging to estimate joint damage in rheumatoid arthritis, and the second study [30], which used hand gesture recognition to interpret everyday activities in medical and educational contexts, our study uses hand and visual cues to determine whether instructional videos are suitable for students with ADHD. Our comparison's validity is increased by this common goal, which highlights how crucial it is to choose models that can accurately depict minute hand movements. An accuracy of 91.48% was found in the initial study [29], which used a hybrid RANet+SVM+QNN model on a short dataset of 600 samples. The accuracy of the second study [30], which used a 1D CNN on a dataset of 2,700 samples, was 89.25%. On the other hand, our suggested model, which was trained on a much larger dataset of 10,600 instructional video frames and was based on VGG16, achieved a significantly higher accuracy of 97.84%. This shows the efficiency and expandability of our system. Furthermore, our system adds usefulness beyond classification because it is particularly made for assessing instructional content related to ADHD, in contrast to previous studies.

Table 5. Enhanced model performance comparison with different dataset categories

Paper ID	Model used	Dataset size	Accuracy (%)	Notes
[29]	RANet+SVM+QNN	600	0.9148% and 0.6478%	Early stopping at epoch 5, limited dataset size.
[30]	1D CNN	2,700	0.8925	Applied to hand gestures used in routines in education and medicine.
Proposed model	VGG16	10,600	97.84	High scalability and precision with a focus on video classification that is ADHD-friendly.

## 5. CONCLUSION AND FUTURE WORK

Students with ADHD may find online learning materials highly challenging to use, especially because they lack supportive components that help in understanding and attention maintenance. One of the main causes of these difficulties is the lack of tools for efficiently evaluating content that is especially suited to the requirements of students with ADHD. A thorough five-phases-system is presented in this work with the goal of filling this gap. The evaluation criteria include instructor hand movement, video length, audio-visual quality, and diversity of information. According to preliminary findings, the hand movement model successfully differentiates between instructional videos that improve learning for students with ADHD and those that do not, achieving an accuracy of 97.84%. This system's creation represents a substantial achievement in the field of educational technology. It gives parents, teachers, and content producers useful tools that let them choose or produce videos that are most suited for students with ADHD. This model's excellent performance highlights the value of systematic review in evaluating online material, especially about the components that are most pertinent to maintaining focus and increasing engagement. Future work could improve phase one by creating an adaptive system that modifies the duration of videos according to students' interest and engagement, guaranteeing a succinct and powerful presentation. By taking lighting, sound, and resolution variations into consideration, audio-visual quality in phases two and three can be enhanced, improving performance in a variety of environments. Phase four might be extended to assess more instructional materials in order to improve the diversity of the content. In phase five, increased attention to small changes in instructors' hand movements can enhance the accuracy of identifying instructional movements. Lastly, using AI-driven student engagement analysis may provide insights for producing more successful ADHD-focused content. When combined, these enhancements would promote a learning environment that is more tailored and flexible.

## FUNDING INFORMATION

Authors state no funding involved.

## AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Alshefaa Emam	✓	✓	✓	✓	✓			✓	✓	✓	✓		✓	
Eman Karam Elsied	✓	✓		✓	✓	✓		✓	✓	✓	✓	✓		
Mai Kamel Galab	✓	✓		✓	✓		✓			✓	✓	✓	✓	

C : Conceptualization

M : Methodology

So : Software

Va : Validation

Fo : Formal analysis

I : Investigation

R : Resources

D : Data Curation

O : Writing - Original Draft

E : Writing - Review & Editing

Vi : Visualization

Su : Supervision

P : Project administration

Fu : Funding acquisition

## CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

## DATA AVAILABILITY

Data available at this link: <https://www.kaggle.com/datasets/alshefaaemam/adhd-online-instructor-aoi-dataset>.




## REFERENCES

- [1] C. Lange and J. Costley, "Improving online video lectures: learning challenges created by media," *International Journal of Educational Technology in Higher Education*, vol. 17, no. 1, p. 16, Dec. 2020, doi: 10.1186/s41239-020-00190-6.
- [2] Q. Liu, N. Zhang, W. Chen, Q. Wang, Y. Yuan, and K. Xie, "Categorizing teachers' gestures in classroom teaching: from the perspective of multiple representations," *Social Semiotics*, vol. 32, no. 2, pp. 184–204, Mar. 2022, doi: 10.1080/10350330.2020.1722368.
- [3] E. Yıldırım Demirdöğen, İ. S. Esin, B. Turan, and O. B. Dursun, "Assessing sustained attention of children with ADHD in a class






- flow video task," *Nordic Journal of Psychiatry*, vol. 76, no. 7, pp. 497–506, Oct. 2022, doi: 10.1080/08039488.2022.2064545.
- [4] A. Yeung, E. Ng, and E. Abi-Jaoude, "TikTok and attention-deficit/hyperactivity disorder: a cross-sectional study of social media content quality," *The Canadian Journal of Psychiatry*, vol. 67, no. 12, pp. 899–906, Dec. 2022, doi: 10.1177/07067437221082854.
  - [5] L. Bell, W. Scharke, V. Reindl, J. Fels, C. Neuschaefer-Rube, and K. Konrad, "Auditory and visual response inhibition in children with bilateral hearing aids and children with ADHD," *Brain Sciences*, vol. 10, no. 5, p. 307, May 2020, doi: 10.3390/brainsci10050307.
  - [6] J. Turner, "What do ADHDers need? Working towards establishing guidelines and more ethical methods for designing for and with the neurodivergent," Thesis, School of Arts and Communication (k3), Malmo University, 2023.
  - [7] B. A. Dabwan, F. A. Olayah, H. T. Halawani, A. M. Mashraqi, Y. A. Abdelrahman, and A. F. Shamsan, "Classification of sign language Using VGG16 and VGG19," in *ISAS 2023 - 7th International Symposium on Innovative Approaches in Smart Technologies, Proceedings*, 2023, pp. 1–6, doi: 10.1109/ISAS60782.2023.10391673.
  - [8] Z. Peng, Z. Yang, J. Xiahou, and T. Xie, "Recognizing teachers' hand gestures for effective non-verbal interaction," *Applied Sciences*, vol. 12, no. 22, p. 11717, Nov. 2022, doi: 10.3390/app122211717.
  - [9] S. dos A. Freitas and A. S. de Andrade Neto, "Gestures in the teaching and learning process: a systematic literature review," *Preprint*, May 2022, doi: 10.1590/SciELOPreprints.4045.
  - [10] D. G. Yohannan *et al.*, "Using the hands for learning anatomy," in *Graphic Medicine, Humanizing Healthcare and Novel Approaches in Anatomical Education. Biomedical Visualization*, vol. 3. Springer, Cham, pp. 205–229, 2023, doi: 10.1007/978-3-031-39035-7\_10.
  - [11] R. B. Church, M. Perry, M. A. Singer, S. W. Cook, and M. W. Alibali, "Teachers' gestures and how they matter," *Topics in Cognitive Science*, vol. 17, no. 3, pp. 545–568, Jul. 2025, doi: 10.1111/tops.12755.
  - [12] Z. Hong, Y. Chen, J. Sun, S. Cai, T. Xiaoyu, and A. Wang, "Audiovisual integration decreases inhibition of return in children with ADHD," *Journal of Attention Disorders*, vol. 28, no. 13, pp. 1684–1696, Apr. 2024, doi: 10.1177/10870547241284867.
  - [13] S. J. Bolinger, D. W. Mucherah, and D. A. M. Markelz, "Teacher knowledge of attention-deficit/hyperactivity disorder and classroom management," *The Journal of Special Education Apprenticeship*, vol. 9, no. 1, Jan. 2020, doi: 10.58729/2167-3454.1098.
  - [14] X. Min, G. Zhai, J. Zhou, M. C. Q. Farias, and A. C. Bovik, "Study of subjective and objective quality assessment of audio-visual signals," *IEEE Transactions on Image Processing*, vol. 29, pp. 6054–6068, 2020, doi: 10.1109/TIP.2020.2988148.
  - [15] H. Wu *et al.*, "Exploring video quality assessment on user generated contents from aesthetic and technical perspectives," in *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, Oct. 2023, pp. 20087–20097, doi: 10.1109/ICCV51070.2023.01843.
  - [16] A. M. A. ghani Abdulghani and G. G. M. Dalveren, "Moving object detection in video with algorithms YOLO and faster R-CNN in different conditions," *European Journal of Science and Technology*, 2022, doi: 10.31590/ejosat.1013049.
  - [17] A. Emam, "ADHD online instructor (AOI) dataset," *Kaggle*. Accessed: Jan 29, 2025. [Online]. Available: <https://www.kaggle.com/datasets/alshefaaemam/adhd-online-instructor-aoi-dataset>
  - [18] D. Ezzat, E. Ahmed, E. Soliman, and A. E. Hassanien, "An interpretable Bayesian deep learning-based approach for sustainable clean energy," *Neural Computing and Applications*, vol. 36, no. 27, pp. 17145–17163, Sep. 2024, doi: 10.1007/s00521-024-10008-5.
  - [19] E. A. Mahareek, E. K. ElSayed, N. M. ElDesouky, and K. A. ElDahshan, "Detecting anomalies in security cameras with 3D-convolutional neural network and convolutional long short-term memory," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 14, no. 1, pp. 993–1004, Feb. 2024, doi: 10.11591/ijece.v14i1.pp993-1004.
  - [20] G. Park, V. K. Chandrasegar, and J. Koh, "Accuracy enhancement of hand gesture recognition using CNN," *IEEE Access*, vol. 11, pp. 26496–26501, 2023, doi: 10.1109/ACCESS.2023.3254537.
  - [21] N. Rai *et al.*, "Applications of deep learning in precision weed management: A review," *Computers and Electronics in Agriculture*, vol. 206, p. 107698, Mar. 2023, doi: 10.1016/j.compag.2023.107698.
  - [22] G. R. S. N. Kumar, R. S. Sankuri, and S. P. K. Karri, "Multi scale aided deep learning model for high F1-score classification of fundus images based diabetic retinopathy and glaucoma," in *2023 International Conference on Computer, Electronics & Electrical Engineering & their Applications (IC2E3)*, Jun. 2023, pp. 1–6, doi: 10.1109/IC2E357697.2023.10262434.
  - [23] G. Zeng, "On the confusion matrix in credit scoring and its analytical properties," *Communications in Statistics - Theory and Methods*, vol. 49, no. 9, pp. 2080–2093, May 2020, doi: 10.1080/03610926.2019.1568485.
  - [24] O. G. Ajayi and J. Ashi, "Effect of varying training epochs of a faster region-based convolutional neural network on the accuracy of an automatic weed classification scheme," *Smart Agricultural Technology*, vol. 3, p. 100128, Feb. 2023, doi: 10.1016/j.atech.2022.100128.
  - [25] W. Li *et al.*, "Machine learning model of ResNet50-ensemble voting for malignant–benign small pulmonary nodule classification on computed tomography images," *Cancers*, vol. 15, no. 22, p. 5417, Nov. 2023, doi: 10.3390/cancers15225417.
  - [26] G. Meena, K. K. Mohbey, and S. Kumar, "Sentiment analysis on images using convolutional neural networks based Inception-V3 transfer learning approach," *International Journal of Information Management Data Insights*, vol. 3, no. 1, p. 100174, Apr. 2023, doi: 10.1016/j.jjimei.2023.100174.
  - [27] S. R. Shah, S. Qadri, H. Bibi, S. M. W. Shah, M. I. Sharif, and F. Marinello, "Comparing inception V3, VGG 16, VGG 19, CNN, and ResNet 50: a case study on early detection of a rice disease," *Agronomy*, vol. 13, no. 6, p. 1633, Jun. 2023, doi: 10.3390/agronomy13061633.
  - [28] S. Padmakala, "Deep learning approaches to gastrointestinal image classification: evaluating VGG19 and EfficientNetB0 models," in *2024 7th International Conference on Devices, Circuits and Systems (ICDCS)*, Apr. 2024, pp. 205–210, doi: 10.1109/ICDCS59278.2024.10560955.
  - [29] R. K. Ahalya, F. M. Almutairi, U. Snehalatha, V. Dhanraj, and S. M. Aslam, "RANet: a custom CNN model and quanvolutional neural network for the automated detection of rheumatoid arthritis in hand thermal images," *Scientific Reports*, vol. 13, no. 1, p. 15638, Sep. 2023, doi: 10.1038/s41598-023-42111-3.
  - [30] N. Al Mudawi *et al.*, "Innovative healthcare solutions: robust hand gesture recognition of daily life routines using 1D CNN," *Frontiers in Bioengineering and Biotechnology*, vol. 12, Jul. 2024, doi: 10.3389/fbioe.2024.1401803.




**BIOGRAPHIES OF AUTHORS**

**Alshefaa Emam**    received the B.Sc. degree in information technology from Misr University for Science and Technology (MUST). She is currently pursuing an M.Sc. degree in computers and artificial intelligence at Benha University. She is a teaching assistant at Zewail City, School of computational sciences and artificial intelligence. Her research interests include deep learning, computer vision, and AI-driven educational systems. She has published research in the field of AI and its applications in educational content analysis. She can be contacted email at: shefaaemam97@gmail.com.



**Eman Karam Elsayed**    is the Dean of the College of Information Technology at Misr University for Science and Technology (MUST). She previously served as the Dean of the School of Computer Science at the Canadian International College. She earned her Bachelor of Science degree in computer science from Cairo University in 1994, followed by a Master's degree in computer science from Cairo University in 1999. She obtained her Ph.D. in computer science from Al-Azhar University in 2005 and was promoted to Professor of Computer Science in 2019. With a strong background in artificial intelligence, Prof. Elsayed has published 65 research papers across various AI domains as of March 2024. She can be contacted via email at: eman.karam@must.edu.eg.



**Mai Kamel Galab**    received her BSc in computer science from the Faculty of Science, Benha University, in 2007. She completed her master's degree in computer science at the Faculty of Computers and Artificial Intelligence, Benha University, in 2015. Dr. Galab earned her Ph.D. in 2022 from the same faculty. Since 2022, she has been a lecturer in computer science. Her research interests are computer vision, image processing (human behavior analysis-video surveillance systems), security (steganography-encryption), and human-computer interaction (HCI). She can be contacted at email: may.gallab@fci.bu.edu.eg.