International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 03 March 2026
Page No: 1362-1371
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0188

# AI-Based Inclusive Assistive And Learning Support Platform For Visually Impaired And Autism Students

*Dr. A. Abdul Azeez Khan [1], Dr. K. Javubar Sathick [2*], Mohamed Aaqil M.H [3], Pooja .K [4]*
*[1,2] Associate Professor, Department Of Computer Applications, B.S Abdur Rahman Crescent Institute Of Science And Technology, Chennai.*
*[3,4] Mca 2nd Year, Department Of Computer Applications, B. S. Abdur Rahman Crescent Institute Of Science And Technology, Chennai.*
***EmailId:*** *abdulazeezkhan@crescent.education[1], javubar@crescent.education[2], mhmohamedaaqil@gmail.com[3], pooj2425@gmail.com[4]*

## Abstract
*Inclusive education necessitates the development of technology solutions that can accommodate diverse learning abilities, especially those of visually impaired individuals and those with Autism Spectrum Disorder (ASD). This paper discusses the design and development of an Artificial Intelligence (AI) enabled platform for assistive and learning support in an inclusive education context. The proposed platform is divided into two main modules: a voice-enabled assistive module for visually impaired individuals, and a visual engagement analytics module for Autism Spectrum Disorder students. The assistive module is enabled by Automatic Speech Recognition (ASR), Optical Character Recognition (OCR), YOLO Object Detection, Text-to-Speech (TTS), and an SOS emergency feature. The learning analytics module incorporates eye-tracking data, feature engineering, random forest classification, SHAP explain ability, and argentic reasoning for analyzing student engagement. The publicly available eye-tracking data and synthetically generated data set have been used to test the classification of student engagement and behavioral pattern analysis. The proposed system architecture has adopted a scalable layered architecture for integrating perception modules, machine learning models, and explainable learning analytics modules. The proposed assistive computing and educational analytics framework can be useful for creating an inclusive learning environment.*

***Keywords:*** *Artificial Intelligence; Assistive Technology; Autism Education; Eye Tracking, Explainable AI; Visual Impairment.*

## 1. Introduction

Disability Inclusive education seeks to ensure that individuals with different abilities receive equal learning opportunities. However, individuals with disabilities, including visually impaired individuals and those with autism spectrum disorder (ASD), tend to face various challenges in accessing learning resources. For instance, visually impaired individuals tend to be limited in their ability to recognize objects, read, and move around in unfamiliar places without guidance. The limitations tend to affect their learning independently, as most learning materials tend to be in formats that require individuals to be visually oriented. In the same way, students with autism have unique behavioral and cognitive characteristics that affect their learning and interaction with their learning environment. Autism spectrum disorder is often linked with social communication differences, sensory processing differences, and attention pattern differences. When autism students are in an educational environment, they are often observed to have different attention patterns and eye behaviors from their non-autism counterparts or students. Understanding their visual attention pattern is significant and could give valuable insights into their engagement and learning behavior. Although there is an increasing availability of assistive technologies, existing assistive technologies that cater to the visually impaired population generally offer only isolated capabilities like screen reading, OCR-based text detection, or object detection applications. They generally do not integrate into an overall system. Similarly, many eye-tracking-based autism research systems generally do not offer overall useful learning analytics but rather focus on clinical diagnosis or

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 03 March 2026
Page No: 1362-1371
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0188

screening. Additionally, many artificial intelligence-based systems that offer learning analytics capabilities generally function as black box models, making it hard for educators to understand the results generated by the system. To overcome the limitations of existing assistive technologies, the proposed research aims to design an overall AI-based inclusive assistive learning support system. It aims to integrate accessibility technologies with intelligent learning analytics. It proposes the integration of real-time perception technologies like Automatic Speech Recognition (ASR), Optical Character Recognition (OCR), and CNN-based object detection with overall learning analytics capabilities like Random Forest classification along with SHAP-based explain ability to offer useful learning analytics results to educators regarding the overall visual engagement behavior of autism students. The novelty of the proposed approach is based on the integration of assistive computing and educational analytics into one architecture. The proposed architecture is intended to provide support for both independent accessibility by visually impaired students and engagement analysis for autism students, so that the effectiveness of AI-based learning technology is increased.

## 2. Related Work

### 2.1. Assistive Technologies For Visual Impairment

Assistive technologies for visually impaired people have mostly focused on the development of accessible interfaces using speech and computer vision. Existing assistive technologies have mostly focused on the development of accessible interfaces using speech and computer vision. They have mostly focused on converting visual information, like text or objects in the environment, into speech for the visually impaired. Saleem et al. [1] proposed a visually impaired assistive system using OCR technology, which can be used for converting text into speech for the visually impaired. Gao et al. [2] proposed techniques for improving the accuracy of the OCR system using image preprocessing techniques like noise filtering and adaptive thresholding. In addition, speech recognition has been used in various assistive technologies. Radford et al. [3] proposed a robust speech recognition system that utilized large-scale weak supervision. This improved the accuracy of speech-based interaction systems. Other studies utilized object detection models that utilized CNNs and YOLO. This helped in recognizing the objects present in the environment. Patel et al. [10] proposed an assistive voice-enabled mobile application that utilized object detection along with audio feedback. Although these studies contribute significantly to the domain of assistive technologies, most studies focus on individual assistive services. However, the integration of various assistive services into a unified system is still in its infancy. This proposed system is motivated by the various studies that utilized individual assistive services. It aims to integrate voice interaction, OCR, object detection, and emergency assistance into a unified system.

### 2.2. Eye-Tracking and Autism Engagement Analysis

Eye-tracking technology was extensively employed to analyze the visual attention patterns of people with autism spectrum disorder (ASD). Researchers have attempted to identify the gaze patterns as a tool for understanding the cognitive engagement and behavioral differences among autism students. Kanhirakadavath and Chandran [4] attempted to analyse the scan path for autism screening using machine learning approaches. Their results indicated that the gaze patterns can identify the behavioral characteristics among people with autism spectrum disorder. In a similar vein, Chetcuti et al. [6] examined the feasibility of employing short-term eye-tracking for autism identification. Their results indicated that even a short-term gaze-tracking session can provide valuable insights into the behavioral characteristics among people with autism. Jeyarani et al. [8] further explored the application of deep learning models for autism identification using the eye-tracking approach. Their results indicated that machine learning models can effectively analyse the complex gaze patterns for effective classification. However, despite the progress that has been made in this area, the majority of the existing studies have concentrated on classification and diagnosis. Little emphasis has been put on the analysis of the patterns of engagement in classroom environments. Moreover, the existing models lack interpretability, which makes it hard for educators to understand the

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 03 March 2026
Page No: 1362-1371
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0188

logic behind the classification results. This study seeks to bridge this gap by focusing on the analysis of engagement and the use of explainable machine learning.

### 2.3.Explainable AI in Education

Explainable Artificial Intelligence (XAI) is becoming more significant in the context of educational systems, as transparency and trust are vital in this area. Machine learning models are often used as black boxes in AI systems, and this is not beneficial for educators.[4] Ribeiro et al. [9] highlighted the significance of interpretable machine learning models in improving trust and usability in AI-based decision systems. The authors proposed models that allow users to understand the contribution of different features to the AI models. Additionally, Antoli et al. [5] used explainable machine learning models in eye-tracking analysis and demonstrated how feature importance is used to understand behavioral aspects. Among the various explainability techniques proposed so far, SHAP is identified as an effective technique that provides both global and local explanations for the models. SHAP values help to determine the impact of various features and are used by researchers and educators to better understand the various aspects affecting the classification outcomes. Explainable AI has been used in various research fields; however, there is limited research available that uses explainability in conjunction with assistive technology and learning analytics. The proposed system is motivated by such research and uses SHAP-based explainability with machine learning models to provide effective engagement analysis in inclusive education systems.

## 3. Methodology

The proposed methodology is used on integrating perception, analytics, and interpretability in terms of an inclusive structured processing pipeline that consists of various layers of functions that work together to facilitate the operation of the proposed inclusive assistive learning support platform. At the highest layer of the proposed architecture is the presentation layer, which is the interface between the system and the user. This layer consists of s voice-based mobile interface enables the user to interact with the system using voice commands. Additionally, the system provides a web-based dashboard that is provided to educators / caretakers. This enables them to upload eye-tracking data view, engagement analytics and download reports. At the next layer is the control and routing layer. [13-16]This layer is responsible for controlling the operations of the system by interpreting the user's input commands and directing them to the appropriate functions. This layer consists of various functions such as the command validation engine, intent recognition module, module routing controller, session manager, etc. At the next layer of the proposed architecture is the assistive processing pipeline and the learning analytics pipeline, both of which are in parallel. At the assistive processing mainly deals with real-time accessibility services for the visually impaired.[7] It compromises a voice processing unit that incorporates adaptive voice capture and automatic speech recognition, an optical character recognition unit that process the captured images using various preprocessing techniques and the tesseract OCR engine and an object detection unit that incorporates CNN and YOLO object detection techniques using the captured video feed from the cameras. The recognized data is converted into audio form using the text-to-speech unit. Additionally, the assistive processing module comprises a safety unit that incorporates GPS location tracking, SOS messages and emergency contact communications. At the same time, the learning analytics module processes the eye-tracking data for autism engagement analysis. It comprises dataset upload and parsing, preprocessing mechanisms that compute various behaving features such as fixation duration, face-object attention ratio, gaze entropy and gaze transition frequency. These features are fed to a machine learning classifications engine, where a random forest classifier examines engagement trends and identifies students based on their attention profiles. To enhance transparency and interpretability, the architecture incorporates an explainability subsystem using SHAP analysis, which determines the contribution of individual features to classification decisions and supports an agentic reasoning component that converts analytical results into understandable educational insights. All processing modules interact with a centralized AI analytical layer, where machine learning models, object detection networks and explainability

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 03 March 2026
Page No: 1362-1371
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0188

algorithms operate together to generate predictions and insights. At the foundation of the architecture lies the data management layer, which stores user profiles, eye-tracking datasets, extracted feature matrices, trained model parameters, classification outputs and generated reports. It also showcase the flow of safety layer for SOS connection contact when triggerd to ensure audio feedback for visually imapired students.

### 3.1.System Architecture

The proposed system has a layered and modular architecture that enables efficient integration of assistive services and learning analytics in a single framework. This layered architecture breaks down users' interaction, data handling, smart analysis, storage, and safety features into separate functional components. This makes the system maintainable, scalable and importantly reliable. The architecture has five layers: Presentation Layer, Application Layer, AI Analytics Layer, Data Layer, and Safety Layer. It is shown in Figure 1

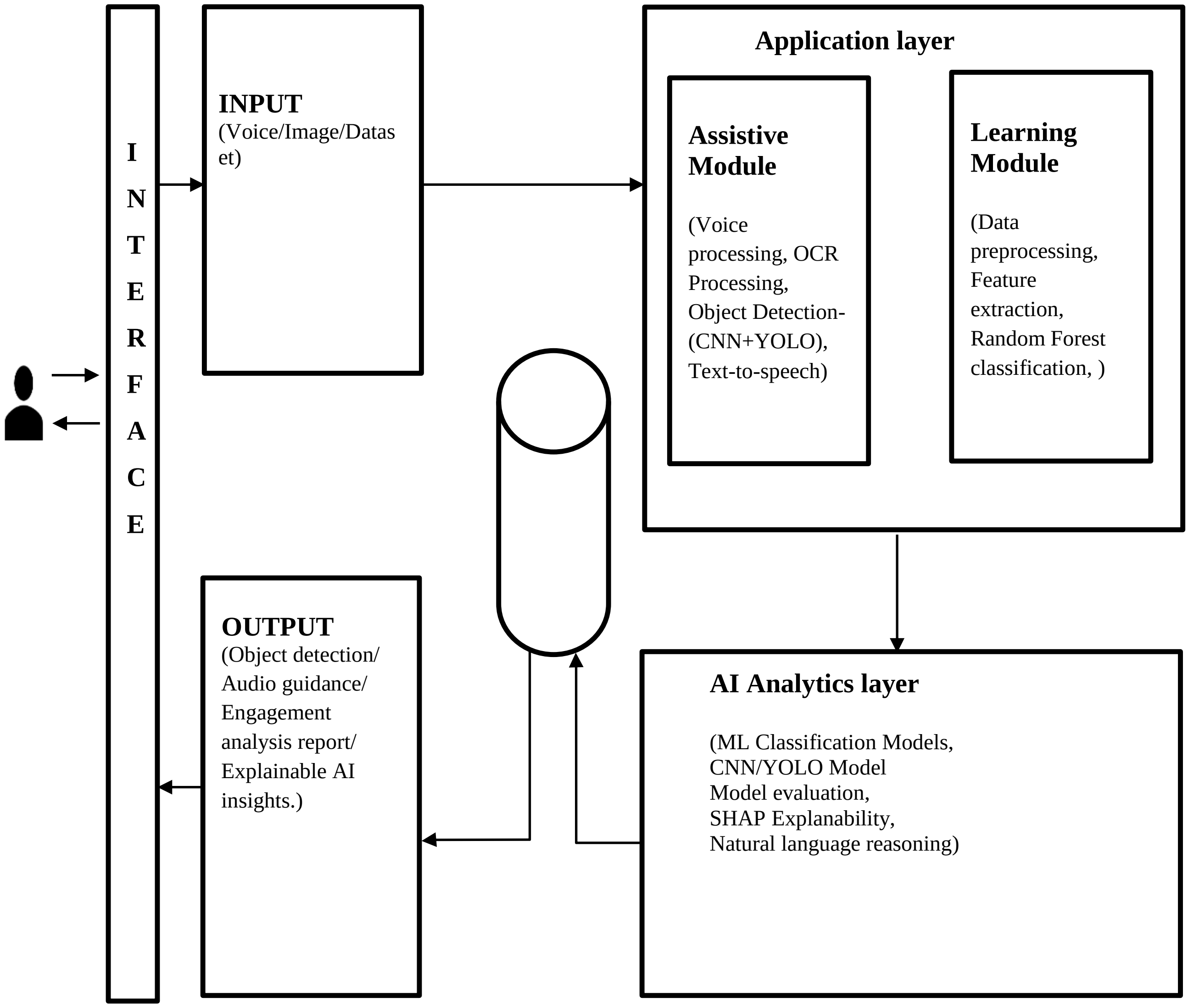


**Figure 1** Architecture of AI- based Assistive learning platform

### 3.1.1. Presentation Layer

The presentation layer acts as the primary interface for users to interact with the platform. It includes a voice-driven interface for a mobile platform that tailored for visually impaired users. This will allow them to effectively interact with the platform using

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 03 March 2026
Page No: 1362-1371
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0188

voice commands. In addition, the platform includes a dashboard for educators/caregivers that will allow them to access engagement analytics and reports generated. It will also feature audio navigation and interactive visualization panels to cater to the needs of all user.

#### 3.1.2. Application Layer

Application layer is responsible for performing all the processing functions of the system. It is responsible for performing voice signal preprocessing, enhancing images for accurate text extraction, performing object detection using YOLO, processing eye-tracking data and feature extraction for engagement analysis and classification.

#### 3.1.3. Ai Analytics Layer

The AI analytics layer acts as the brain of the entire system. The AI analytics layer includes a random forest classifier for the classification of autism engagement, CNN models for object detection, a SHAP engine for explainability and an agentic reasoning module for converting analytical results to useful insights for educators.

#### 3.1.4. Data Layer

The data layer is responsible for the management and storage of the system's information, which is essential for processing and analysis. The information includes user profile, eye-tracking gaze data, extracted features, classification results and generated reports.

#### 3.1.5. Safety Layer

This is ensured by the safety layer, which provides protection to the user by integrating GPS-based location tracking and SOS emergency contacts when triggered, ensuring audio feedback for visually impaired individuals.[25-27] The following figure 1 shows how the user's different type of inputs flows through the application layer and executed according to the inputs (either goes to assistive or learning module). Then the execution moves to Ai and analytics layer Where the ML and CNN model evaluate and classify the data and sends the data to the data layer where it stores the result and report before sending it to the user.

### 3.2.Random Forest Classification

Random forest is an ensemble model consisting of multiple decision trees.
Let training dataset:

$$D = \{(x_i, y_i)\}_{i=1}^{n} \qquad (1)$$

Here in formula 1 $\boldsymbol{D}$ represents the training dataset, $\boldsymbol{x_i}$ represents the feature vector of the i-th sample, $\boldsymbol{y_i}$ = class label of the i-th sample, $\boldsymbol{i}$ = index of the sample, **n** = total number of samples in the datasetEach tree is trained using bootstrap sampling. Prediction function:

$$\hat{y} = mode\{T_{1(x)}, T_{2(x)}, \ldots, T_{k(x)}\} \qquad (2)$$

Here in formula 2 $\hat{y}$ represents the final predicted output of random forest model, Tk indicates the kth decision tree, Mode () represents majority voting(frequency) and $\boldsymbol{x}$ denotes input feature vector
Gini impurity used for split:

$$G = 1 - \sum_{\{i=1\}}^{c} p_i^2 \qquad (3)$$

Here in formula 3 the $\boldsymbol{G}$ is Gini impurity used to measure the quality of a split in a decision tree, $\boldsymbol{pi}$ represents probability of samples belonging to class i, $\boldsymbol{c}$ represents the total number of classes, $\sum$ represents the summation of squared class probabilities.
Gini impurity measures how mixed the classes are in a node. A lower value of $G$ indicates a purer node, which results in a better split in the decision tree.
Feature importance:

$$FI_j = \sum_{splits} \Delta Gini_j \qquad (4)$$

Here in formula 4 $\boldsymbol{FIj}$ represents the importance score of the $j^{th}$ feature in the Random Forest model. The term $\boldsymbol{\Delta Gini_j}$ represents the decrease in Gini impurity produced by splits using the $j^{th}$ feature across the decision trees. The summation over splits indicates that the total contribution of that feature is calculated by adding the impurity reduction from all nodes where the feature is used for splitting. [28-30]A higher value of $FI_j$ indicates that the feature contributes more significantly to improving the model's classification performance. It is shown in Table 1.

### Cnn-Based Object Detection (Yolo)

CNN extracts spatial features using convolution:

$$f(x) = (x * w) + b \qquad (5)$$

In formula 5 $\boldsymbol{x}$ represent the input image, $\boldsymbol{wr}$ epresents the filter kernal, $\boldsymbol{b}$ is the bias of system.

**Table 1** Feature Set Used

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 03 March 2026
Page No: 1362-1371
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0188

| Feature | Description |
|---|---|
| Average Fixation Duration | Mean gaze fixation time |
| Face Ratio | Face attention /total fixation |
| Object Ratio | Object fixation proportion |
| Gaze Entropy | Attention variability |
| Transition Count | ROI switching frequency |

### Activation (Relu):

$$f(x) = \max(0, x) \tag{6}$$

Here in formula 6, $\boldsymbol{f(x)}$ stands for the output of the activation function and $\boldsymbol{x}$ is just the input fed into the neuron. The rule's pretty simple, if the input's positive, you get that value back; if not , you get zero. people usually call this the Rectified Linear unit, or ReLu. You'll see it a lot in Convolutional Neural Network(CNNs) because it's fast and it gives the network that crucial non-linearity

### Loss Function In Detection:

$$Loss = \lambda_{coord} \sum (x - \hat{x})^2 + \lambda_{conf} \sum (C - \hat{C})^2 \tag{7}$$

In formula 7 the ***Loss*** indicates the total loss used to train the object detection model. The term $\boldsymbol{x}$ represents the actual bounding box coordinates of the object, while $\boldsymbol{x}$ denotes the predicted bound box helps generated by the model. C indicates the actual confidence score indicating whether an object is present, and $\boldsymbol{C}$ is confidence loss components, this loss function helps in the learning of the model for accurate object locating and detection. It is shown in Table 2

YOLO predicts bounding box coordinates and confidence scores.

### Shap Explainability

$$\emptyset i = \sum_{\{s \subseteq f \setminus \{i\}\}} \frac{|s|!\,(|f| - |s| - 1)!}{|f|!} [f(s \cup \{i\}) - f(s)] \tag{8}$$

Here in formula 8, $\boldsymbol{f}$ is feature set, $s$ is the subset of features and $\boldsymbol{\emptyset i}$ is the contribution of feature i.

**Table 2** Dataset Summary

| Dataset | Type | Samples | Purpose |
|---|---|---|---|
| Dataset A | Public Eye Tracking | 420 | Validation |
| Dataset B | Public ASD Dataset | 350 | Testing |
| Synthetic Dataset | Generated | 300 | Pattern simulation |

## 4. Experimental Results

The performance of the proposed module for learning analytics was tested using the following datasets:

Dataset A – Public Eye-Tracking Dataset 1 – 420 samples

Dataset B – Public Eye-Tracking Dataset 2 – 350 samples. It is shown in Table 3

**Table 3** Dataset Distribution Table

| **Dataset** | **Samples** | **Engagement Classes** | **Usage** |
|---|---|---|---|
| Public Dataset A | 420 | 4 | Training + Validation |
| Public Dataset B | 350 | 4 | Testing |
| Synthetic Dataset | 300 | 4 | Robustness Evaluation |
| **Total** | **1070** | **4 Classes** | — |

Synthetic Dataset – Generated – 300 samples

For the synthetic dataset, the proposed approach for generating synthetic data was based on the generation of realistic eye-tracking behaviors with the following for engagement profiles:

1. Face-Oriented Attention
2. Object-Oriented Attention
3. High Variability Attention

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 03 March 2026
Page No: 1362-1371
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0188

4. Sustained Focus Attention

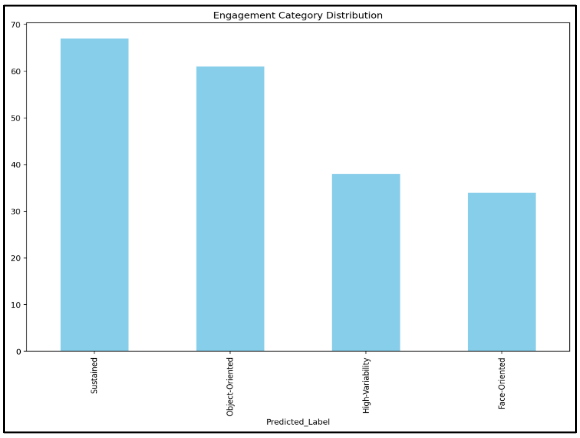


**Figure 2 Engagement classification Distribution**

The figure 2 displays the distribution of participants spread among the four classes (Sustained focus, Object oriented, High-Variability and sustained focus).

The generation of synthetic data was based on the variation of the features using probabilistic modelling:

$Entropy = -\sum p_i \log(p_i)$

Here pi represents fixation probability within defined Regions if Interest (ROIs) the synthetic dataset ensured balanced class representation and controlled feature variation to test model robustness.

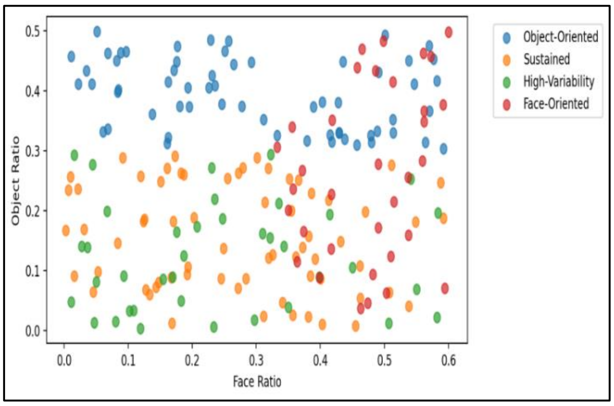


**Figure 3 Scatter plot of gaze coordinates between object and face ratio**

The figure 3 displays the gaze feature spread across to decision making attribute “Object ratio” and “Face ratio”. The scatter plots denotes the students gaze behavior and plot them in their respective classes and represented as:

- Blue dot – Object - Oriented
- Orange dot – Sustained focus
- Green dot – High - variability
- Red dot - Face oriented

## 4.1.Assistive Module Evaluation

The assistive module was tested under controlled real-world scenarios such as:

- Quiet indoor environment
- Moderately noisy classroom
- Artificial background noise simulation
- Variable lighting conditions

Speech recognition was tested by sending 200 voice commands under different conditions.[17-20] The performance of OCR was tested with 150 printed images under varying illumination levels. The performance of the object detection was tested with live camera feed and 20 different types of objects.

CNN Object Detection Analysis

Object detection performance was evaluated using standard metrics:

$$Precision = TP/(TP+FP)$$

Here in 9, The precision denotes the proportion of correctly predicted positives observed among all predicted observations. The term TP (True positive) denotes the numbers of correctly found positive instance and FP (False positive) denotes the incorrectly found positives

$$Recall = TP/(TP+FN)$$

Here in 10, recall denotes the proportion of actual positive instances are correctly identified by the model.

$$mAP = 1/N \sum APi$$

Here in 11, mAP represent the average performance of the object detection model across all the classes, APi indicates the average precision of the ith class, while N indicates the total no of classes used in model evaluation for object detection. It is shown in Table 4

**Table 4: Performance Evaluation Table**

| Module | Metric | Outcome |
|---|---|---|
| Speech Recognition | Accuracy | 92% |
| OCR | Accuracy | 90% |
| Object Detection | Accuracy | 88% |

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 03 March 2026
Page No: 1362-1371
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0188

| | | |
|---|---|---|
| Engagement Classifications | Accuracy | 91% |
| Precision (RF) | Learning Module | 91% |
| Recall (RF) | Learning Module | 90% |
| F1-Score (RF) | Learning Module | 89% |
| Response Time | Assistive Module | <2 seconds |

## 5. Discussion

From the results obtained in the experiments, it is evident that the integration of assistive technology and learning analytics within the same framework using AI has the potential for improving the level of accessibility and the generation of educational insights.[12]The accuracy of the assistive technology module was found to be high under controlled and moderately challenging environments. The accuracy of the speech recognition system was found to be consistent at over 90%, expect for some degradation in the presence of heavy noise. The accuracy of the OCR system was affected significantly by the quality of lighting. The object detection module was successful in the identification of objects, which validates the effectiveness of the CNN-based YOLO architecture.[21-24] The low latency (< 2 seconds) is also significant for real-time interaction with visually impaired users relying on sound alone. In the case of the autism learning analytics module, the Random Forest classifier had an overall accuracy of 91% with the incorporation of real and synthetic datasets. The precision and recall values are well-balanced, which is significant for effective generalization across engagement types. The incorporation of synthetic datasets is significant in robustness testing. Significantly, the inclusion of SHAP-based explainability results in a considerable improvement in transparency. Unlike the conventional black box model, the proposed model's results provide interpretable insights that educators can trust. The feature contribution results for the model prove that fixation duration and face ratio are the primary factors for the classification of engagement. Furthermore, the inclusion of the layered architecture results in a considerable improvement in the scalability of the model. Each component of the model, including assistive perception, analytics and explainability, operated independently. This results in the model's ease of upgrade. Despite the environmental factors affecting the model's perception component, such as lighting conditions and noise levels, the model's results operate within acceptable operating limits. The results prove the feasibility of the model for the deployment of an inclusive AI model.

## Conclusion

This research proposes an inclusive assistive and learning support platform that utilizes artificial intelligence to overcome the challenges that visually impaired people face while providing a learning analysis for autism students. The platform utilizes various artificial intelligence techniques such as voice interaction, Optical Character Recognition, real-time object detection, eye tracking data analysis, Random Forest for user engagement classification and SHAP for explainable artificial intelligence. The assistive component will allow visually impaired people to navigate their surroundings using voice commands, text detection and object detection while providing real-time audio feedback and safety assistance. At the same time, the learning component will analyse the eye tracking data to identify the learning characteristics for autism students. The experimental evaluation shows that the proposed platform is successful in providing reliable performance for both functional modules. The assistive module is effective in the presence of a controlled and moderately noisy environment, where high accuracy is obtained in speech recognition, OCR, object detection and low latency is ensured for real-time interaction. The learning analytics module obtains high classification accuracy in the classification engagement profiles using the proposed model of Random Forest and the addition of SHAP provides clear and transparent results for the factors affecting the model. These results are useful for the educator in understanding the behavior of the students and developing affective teaching strategies. Additionally, the modular layered architecture of the proposed platform is useful for scaling the system and incorporating future enhancements such as the addition of deep learning-based gaze analysis, multilingual support, offline analysis and large-scale deployment in inclusive educational environments.

Further this proposed model for an AI- based assistive learning support platform help other in developing more easily accessible, student and educator friendly platforms or applications in future. The future enhancement for the research would be collecting real time eye-tracking gaze datasets from autistic students. Developing and streamlining the platform as open source for user to customize the interaction as per user preference and help the visually impaired student to customize the platform their unique needs and accessibility

Therefore, these enhancement in future help the students with assistive needs and educators to help the students furthermore and make learning more accessible. As this research took inspiration from the previous research and journals, we believe that this AI- Based inclusive assistive, learning support platform-oriented research will help the innovators to build more accessible learning platform for more assistive and special need children and make an impact on future education.

## References

[1]. Shahid Saleem, Muhammad Awais Khan, and Amir Abbas, “Smart assistive system for visually impaired using OCR and speech synthesis,” IEEE Access, vol. 9, pp. 1–15, 2021.

[2]. Qing Gao et al., “VI-OCR: Visually Impaired Optical Character Recognition,” Scientific Reports, 2025.

[3]. Alec Radford et al., “Robust speech recognition via large-scale weak supervision,” arXiv:2212.04356, 2023.

[4]. M. R. Kanhirakadavath and M. S. Mohan Chandran, “Investigation of Eye-Tracking Scan Path as a Biomarker,” MDPI Diagnostics, vol. 12, no. 7, pp. 1–18, 2022.

[5]. Jordi Antoli et al., “Explainable Machine Learning and Eye Tracking,” Frontiers in Neuroscience, vol. 19, 2025.

[6]. Luke Chetcuti et al., “Feasibility of a Two-Minute Eye-Tracking Protocol,” Scientific Reports, vol. 14, 2024.

[7]. Carlos Fernández-Llamas et al., “Technology-Supported Inclusive Education,” Frontiers in Education, vol. 10, 2025.

[8]. R. Jeyarani et al., “Applying Eye Tracking with Deep Learning,” MDPI Data, vol. 8, no. 3, 2023.

[9]. Marco Tulio Ribeiro et al., “Explainable Artificial Intelligence in Educational Applications,” Applied Sciences, vol. 15, no. 4, 2025.

[10]. Rohan Patel, Karan Mehta, and Dhruv Shah, “Voice-enabled mobile assistive application,” Proc. IEEE ICAIAT, 2025.

[11]. M. Ahmed et al., “Eye Tracking-Based Diagnosis of Autism,” MDPI Electronics, 2022.

[12]. S. Alsharif et al., “Deep Learning in Eye-Tracking Systems,” Frontiers in Medicine, 2024.

[13]. Zhao Z. et al., “Classification of Children with Autism Using Eye-Tracking Data,” Journal of Medical Internet Research, 2021.

[14]. K. Hudry et al., “Two-Minute Eye-Tracking Assessment for Autism,” Springer Nature, 2025.

[15]. A. Setu et al., “Eye-Tracking-Based Autism Screening,” Journal of Clinical Medicine, 2025.

[16]. G. Jocher et al., "YOLOv5: Real-Time Object Detection for Edge and Embedded Devices," arXiv:2104.02767, 2021.

[17]. S. Prokhorenkova et al., "CatBoost: Unbiased Boosting with Categorical Features," IEEE Transactions on Knowledge and Data Engineering, vol. 34, no.

[18]. Lundberg S. and Lee S., “A Unified Approach to Interpreting Model Predictions,” NeurIPS, 2017.

[19]. Goodfellow I., Bengio Y., and Courville A., Deep Learning, MIT Press, 2016.

[20]. Bishop C., Pattern Recognition and Machine Learning, Springer, 2006.

[21]. Salesky E. et al., “Direct Speech-to-Speech Translation,” ACL, 2021.

[22]. Bansal S. et al., “Speech-to-Speech Translation for Low-Resource Languages,” IEEE Access, 2024.

[23]. Li X. et al., “End-to-End Multimodal Speech

Translation," IEEE Transactions on Audio, 2025.
[24]. Zhang B. et al., "Neural Speech Translation with Large Language Models," INTERSPEECH, 2023.
[25]. Tesseract OCR Engine Documentation, Google, 2023.
[26]. Y. Liu et al., "Vision-Based Assistive Navigation System for Visually Impaired People Using Deep Learning," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 30, pp. 2158-2168, 2022.
[27]. A. Alsharif et al., "Deep Learning-Based Eye Tracking Analysis for Autism Spectrum Disorder Detection," *Sensors*, vol. 22, no. 18, pp. 1-16, 2022.
[28]. M. Hossain et al., "An Explainable Artificial Intelligence Framework for Educational Data Analytics," *IEEE Access*, vol. 10, pp. 118245-118258, 2022.
[29]. J. Kim et al., "Real-Time Object Detection for Assistive Technologies Using Lightweight CNN Models," *Computer Vision and Image Understanding*, vol. 226, pp. 1-12, 2023.
[30]. R. Sharma et al., "Eye-Gaze Based Learning Analytics for Autism Spectrum Disorder Using Machine Learning Techniques," *IEEE Transactions on Learning Technologies*, vol. 17, no. 2, pp. 250-262, 2024.nce in Educational Applications," *Applied Sciens*, vol. 15, no. 4
(ICAIAT), pp. 1–8, 2025.