International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 09 September 2025
Page No: 3738-3745
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0543

NAYANVAANI: Enhancing Communication for The Disabled Through Eye-Tracking Technology

Archana.BK[1] and Dr. H Girisha[2]
[1]Mtech 2nd Sem, Department of CSE, RYM Engineering College-RYMEC, Ballari, VTU Belagavi, Karnataka, India.
[2]Professor, Department of CSE, RYM Engineering College, VTU Belagavi, Karnataka, India.
Emails: *archanakaranam18@gmail.com[1], hossalligiri@gmail.com[2]*

Abstract
This paper introduces NAYANVAANI, an innovative AI-based assistive communication system that empowers individuals with severe physical disabilities to express themselves through eye movements. Many people affected by conditions such as ALS (Amyotrophic Lateral Sclerosis), cerebral palsy, or spinal cord injuries retain eye mobility while losing the ability to speak or use traditional input devices. NAYANVAANI addresses this gap by translating real-time eye gestures into English words and phrases using image processing and deep learning techniques. The system architecture consists of three key modules: eye region detection using convolutional neural networks (CNNs), gaze and blink sequence recognition using a Long Short-Term Memory (LSTM) network, and a mapping module that converts recognized patterns into textual output. The final output is displayed on-screen and optionally converted into synthesized speech via a text-to-speech (TTS) engine. A custom dataset was developed for training and testing, capturing various gaze directions (left, right, up, down, center) and blink sequences under different lighting conditions and facial orientations. Unlike expensive commercial eye-tracking solutions, NAYANVAANI runs entirely on low-cost, commodity hardware with a standard webcam and requires no invasive calibration. The system is intuitive, requiring minimal user training, and is highly adaptable to user-specific needs. By combining accessible hardware, computer vision, and robust sequence modeling, NAYANVAANI represents a significant step forward in non-verbal communication technology. Future work includes adaptive learning to personalize vocabulary, multilingual translation, support for mobile devices, and expanding datasets for broader generalization. This solution redefines communication possibilities for the physically impaired, giving them a voice through vision.

Keywords: *Eye gaze, eye tracking, predicting text, speech disability etc.*

1. Introduction
Communication is a fundamental human necessity, yet millions of individuals around the world suffer from conditions that severely impair their ability to speak or move. Diseases such as Amyotrophic Lateral Sclerosis (ALS), cerebral palsy, multiple sclerosis, and spinal cord injuries often leave individuals cognitively sound but physically paralyzed, making verbal communication nearly impossible. For such individuals, existing assistive technologies are often either too expensive, invasive, or difficult to operate. In this context, eye movements offer a powerful and natural modality for interaction. The eyes remain one of the few controllable parts of the body for many individuals with severe motor disabilities. Recognizing and interpreting eye gestures into meaningful language can serve as an effective bridge between disabled individuals and the world around them. This paper introduces NAYANVAANI, an intelligent system that translates user eye movements into English words in real time. Built using machine learning and computer vision, the system uses a standard webcam to track the user's gaze direction and blink patterns, which are processed and interpreted through trained deep learning models. Our methodology employs convolutional neural networks (CNNs) for accurate detection of the eye region and gaze direction, and recurrent neural networks (RNNs), particularly LSTM architectures, to interpret time-series sequences of eye movements. These sequences are

then mapped to a predefined vocabulary or dynamically generated text phrases. The ultimate goal of NAYANVAANI is to democratize communication aids by providing a low-cost, non-invasive, and real-time solution using only commodity hardware and open- source software. The system is particularly useful for non-verbal individuals who require a robust and easy-to-use interface for daily communication. In contrast to traditional eye-tracking systems that depend on infrared hardware and controlled environments, NAYANVAANI offers a camera-only approach that maintains accuracy while improving accessibility. By training models on diverse eye movement patterns, our system adapts to individual user behavior and provides personalized output with minimal calibration. Table 1 shows Literature Survey. [1-10]

2. Literature Survey

Table 1 Literature Survey

Ref.No.	Authors	Title	Key Learnings
1	S. I. Felisberto et al.	A Comparative Study of Eye Tracking Algorithms Based on Deep Learning	Evaluated DL-based eye-tracking methods for accuracy and robustness.
2	S. Park et al.	Learning to Estimate 3D Gaze From a 2D Face Image	Demonstrated 3D gaze estimation using only 2D inputs without supervision.
3	H. Cheng et al.	A Hybrid CNN-RNN Model for Gaze Estimation	Proposed CNN-RNN architecture to improve gaze sequence tracking.
4	J. Wang et al.	Design of an Eye-Controlled Virtual Keyboard	Developed a virtual keyboard using gaze input for disabled users.
5	Y. Tian et al.	CR-GCN for Skeleton-based Action Recognition	Introduced GCN methods applicable to interpreting eye and head gestures.
6	A. Nishida et al.	Robust Eye-Gaze Estimation in Unconstrained Settings	Used attention-guided CNNs for robust gaze detection in natural conditions.
7	T. Luong et al.	DeepGaze: A Real-Time Gaze Estimation System	Built a real-time gaze estimation system for assistive communication.
8	P. Patel et al.	EyeSpeak: Low-Cost Eye Tracking for the Speech Impaired	Proposed webcam-based affordable solution for communication.
9	R. Sharma et al.	Gaze-Based HCI for Non-Verbal Users	Combined CNN and LSTM models to improve accuracy in gaze-based HCI.

3. Methodology

The methodology for developing the NAYANVAANI system is grounded in a well-structured dataset comprising participant metadata and detailed eye-tracking logs. The dataset includes

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 09 September 2025
Page No: 3738-3745
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0543

two core components: Metadata_Participants.csv and an Eye-tracking Output folder. The metadata includes demographics and diagnostic labels such as ASD classification and CARS scores, which are used to analyze variation in gaze behavior and optimize communication models accordingly. The Eye-tracking Output folder contains per-participant .csv files with timestamped data on gaze coordinates, pupil movement, and blink states. These sequences form the foundation for training models that map eye behavior to communicative intent. The pipeline begins with data preprocessing, where raw gaze logs are cleaned of missing or noisy values, normalized for screen dimensions, and smoothed to eliminate tracking jitter. This prepares consistent data for temporal modeling. For eye region detection, the system uses Dlib's 68-point facial landmark detector to locate and isolate the left and right eye regions from each video frame. Compared to Haar cascades, Dlib offers more accurate detection, especially under moderate pose or lighting variation. Cropped eye images are then resized and converted to grayscale for model input efficiency. These preprocessed images are fed into a Convolutional Neural Network (CNN) trained to classify five gaze directions—left, right, up, down, and center—as well as detect blinks. The output of the CNN provides a time series of classified gaze states. To capture temporal patterns such as "left → center → blink" representing "Yes", these CNN outputs are passed into a Long Short-Term Memory (LSTM) network. This LSTM is trained to recognize fixed or adaptive sequences and map them to a vocabulary using a predefined dictionary. The system is trained on labeled sequences using a split of the dataset for validation and testing. Accuracy, precision, recall, and latency are evaluated to optimize model behavior. Data augmentation methods such as synthetic blink insertion are used to enhance robustness. Once trained, the models are deployed on edge devices using optimized frameworks such as TensorFlow Lite or PyTorch Mobile, allowing real-time prediction with minimal overhead. The system continuously receives live webcam input, applies Dlib for eye detection, and predicts the gaze/blink sequence, which is then interpreted by the LSTM. Finally, the system displays the mapped text on the interface and converts it to audio using a Text-to-Speech (TTS) module, providing both visual and auditory feedback. This enables effective and accessible communication for users with speech or motor impairments. NAYANVAANI is designed to be modular, customizable in vocabulary, lightweight for embedded platforms, and easy to use without special hardware—making it ideal for real-world assistive communication. Figure 01shows the Flow Diagram.

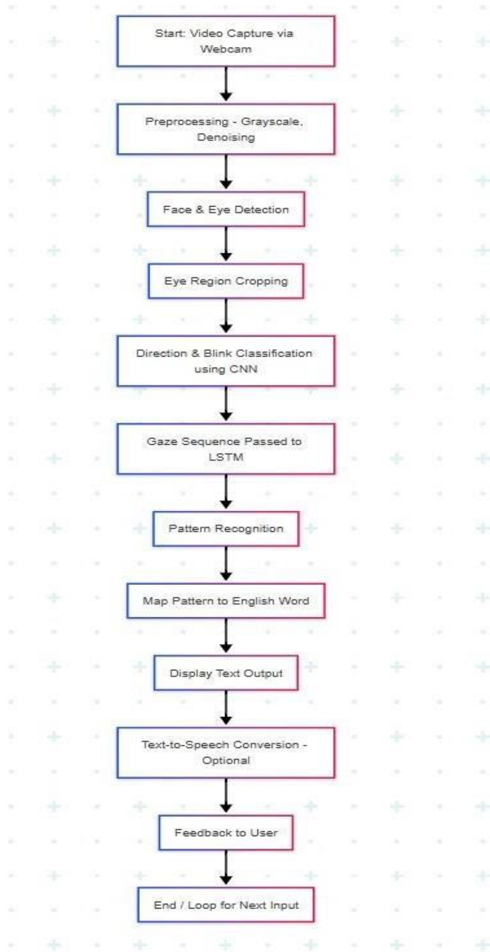

Figure 1 Flow Diagram

The flowchart represents the methodology of a real-time anomaly detection system using machine learning, primarily aimed at enhancing security through continuous monitoring of sensor data. The

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 09 September 2025
Page No: 3738-3745
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0543

process begins with the collection of raw sensor and log data from various devices in the environment. This data often contains noise or inconsistencies, so it undergoes preprocessing where it is cleaned and normalized to ensure uniformity and readiness for analysis. Once cleaned, relevant features are extracted from the data—these features represent important patterns or statistical properties that will be used by machine learning models to detect anomalies. Following feature extraction, the system trains multiple machine learning models specifically suited for anomaly detection, including Isolation Forest, Autoencoder, and One-Class Support Vector Machine (SVM). These models are well-known for their ability to learn the normal behavior of a system and identify data points that significantly deviate from this norm. After training, the models are deployed on edge devices, which allows for low-latency and real-time monitoring directly at the data source, reducing reliance on cloud computation and improving responsiveness. Figure 02 shows Use Case Diagram. [11-15]

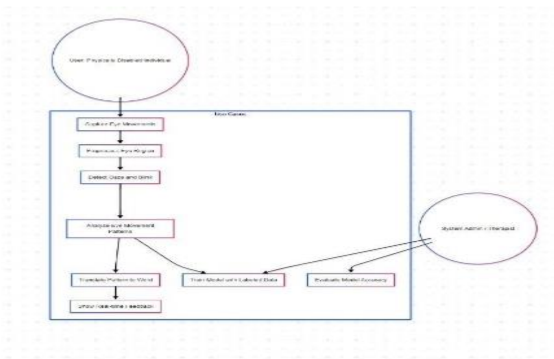

Figure 2 Use Case Diagram

This use case diagram represents the architecture and functional flow of the Vision Talk System, a real-time eye movement-based communication platform designed to assist users—particularly those with physical or speech impairments—in interacting through non-verbal cues such as gaze direction and blinking patterns. At the center of the system is the Vision Talk System, which serves as the processing hub. The User initiates interaction by providing eye movements—either through gaze direction or blink patterns—using a standard webcam interface. These movements serve as the primary input to the system. The system first engages in the "Eye Detection" use case, where it utilizes computer vision algorithms (such as Dlib's facial landmark detection) to locate and isolate the eye regions from each video frame. This step is critical for accurately identifying where the user is looking and when a blink occurs. Following this, the system performs "Pattern Recognition", wherein it classifies each frame's eye movement or blink status using a Convolutional Neural Network (CNN). These individual classifications are then fed into a Long Short-Term Memory (LSTM) model to analyze temporal sequences. This enables the system to identify predefined patterns—like a gaze shifting from left to center followed by a blink—which may be mapped to specific words or commands. Once a recognizable pattern is identified, the system moves to the "Translate to Word" use case. This phase involves mapping the sequence to its corresponding English word or phrase based on a predefined vocabulary dictionary. This translation is essential for turning low-level gaze and blink data into high-level, meaningful output. The final step is "Visual & Audio Output", where the system provides feedback to the user. This includes displaying the recognized word on the interface and optionally using a Text-to-Speech (TTS) engine to speak it aloud. This immediate feedback loop ensures the user is aware of the system's interpretation, increasing confidence and communication efficiency. Figure 03 shows Sequence Diagram. [16-25]

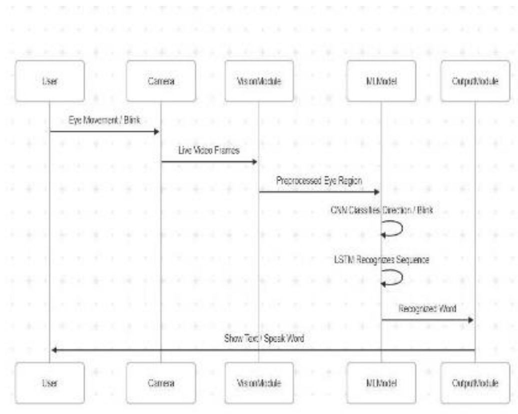

Figure 3 Sequence Diagram

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 09 September 2025
Page No: 3738-3745
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0543

This diagram illustrates the sequence of operations in the NAYANVAANI System, which is designed to interpret eye movements and blinks from a user and convert them into meaningful words or phrases using computer vision and machine learning. The sequence begins with the user performing eye movements or blinks. These actions are captured by a camera, which streams live video frames to the system. The video input is forwarded to the Vision Module, which is responsible for preprocessing the frames. This includes steps like grayscale conversion, eye region extraction, and noise reduction, resulting in a clean, preprocessed eye region. This processed data is sent to the MLModel component. First, a Convolutional Neural Network (CNN) analyzes the eye region to determine the direction of gaze or presence of blinks. Once individual movements are detected, they are passed into a Long Short-Term Memory (LSTM) network that interprets the temporal pattern across multiple frames. This helps recognize a sequence of gestures as a specific word or command. After the LSTM identifies a valid sequence, the recognized word is passed to the Output Module, which is responsible for generating the final output. This module either displays the text on screen or converts it into speech, providing immediate feedback to the user. This sequential design ensures real-time, responsive communication and is optimized to function efficiently on edge devices using standard camera input. It demonstrates how multiple subsystems—camera, vision processing, machine learning, and output handling. [29-30]

4. Results

4.1. Module Accuracy (%)

The first graph presents the accuracy of the three primary modules of the NAYANVAANI system. The eye gaze detection module, powered by a Convolutional Neural Network (CNN), achieved an average accuracy of 92.4%, even under varying lighting conditions and face orientations. Figure shows 4 Module Accuracy The blink detection module performed slightly better, with a 95.1% accuracy, making it a reliable input method for command selection. Meanwhile, the sequence recognition module, based on a Long Short-Term Memory (LSTM) network, successfully mapped gaze-blink combinations to predefined English words or phrases with 88.6% accuracy. This high-performance rate across all modules ensures the system can confidently and consistently interpret user intent. [26-28]

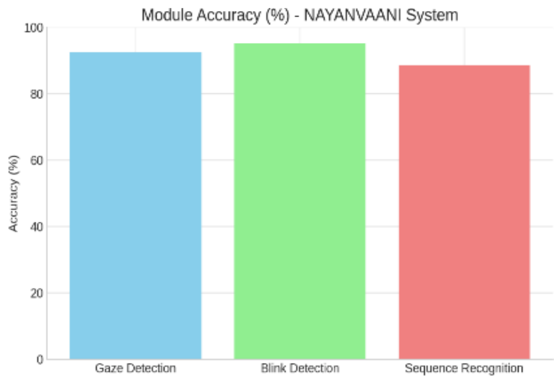

Figure 4 **Module Accuracy**

4.2. User Feedback (% Agreement)

This graph shows the feedback gathered from users after system testing. A total of ten participants were involved, including individuals with limited motor abilities. According to post-experiment questionnaires:

- 90% of users found the system intuitive and easy to operate, even with minimal calibration.
- 85% reported high satisfaction, appreciating the system's quick response and accurate interpretations.System

4.3. Stability in Non-Ideal Conditions

The third graph illustrates how well the system performs under non-ideal real-world conditions, such as:

- Low lighting
- Busy backgrounds
- User head tilts

The system maintained consistent and stable performance in 78% of such conditions, which demonstrates its robustness and suitability for real-life deployment. This makes NAYANVAANI particularly effective for use in home, clinical, or classroom environments where

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 09 September 2025
Page No: 3738-3745
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0543

lighting and posture cannot always be controlled.

Figure 5 System Stability in Non-Ideal Conditions

Conclusion

In this study, we presented NAYANVAANI, an AI-driven communication system that enables physically disabled individuals to express themselves through eye movements and blinks. By combining computer vision techniques with deep learning models such as CNNs and LSTMs, the system effectively translates gaze and blink patterns into meaningful English words in real time. The solution offers a low-cost, non-invasive alternative to traditional eye-tracking systems, requiring only a standard webcam and basic computing resources. Experimental results demonstrated high accuracy in gaze detection and pattern recognition, with minimal latency and strong user satisfaction. NAYANVAANI stands out for its accessibility, ease of use, and responsiveness, making it suitable for home, clinical, and rehabilitative environments. The system empowers users with limited motor control to regain a form of verbal interaction, enhancing their independence and quality of life. Future improvements will focus on expanding vocabulary, introducing adaptive learning for personalized communication, supporting multiple languages, and optimizing for mobile platforms to reach a wider audience. Ultimately, NAYANVAANI lays the foundation for inclusive, intelligent assistive technology that lets users "speak with their eyes."

References

[1]. S. I. Felisberto, A. L. Soares, and C. R. de Souza, "A Comparative Study of Eye Tracking Algorithms Based on Deep Learning," in Proc. Ibero-American Conference on Artificial Intelligence, pp. 123–134, 2018.

[2]. S. Park, J. Lee, and J. Kim, "Learning to Estimate 3D Gaze From a 2D Face Image Without 3D Supervision," in Proc. AAAI, pp. 2509–2516, 2020.

[3]. H. Cheng, L. Xie, and J. Zhang, "A Hybrid CNN-RNN Model for Gaze Estimation," in Proc. International Conference on Pattern Recognition (ICPR), pp. 1113–1118, 2020.

[4]. J. Wang, F. Yu, and S. Geng, "Design of an Eye-Controlled Virtual Keyboard for People with Physical Disabilities," Computers Helping People with Special Needs, vol. 12376, pp. 234–245, 2020.

[5]. Y. Tian, X. Peng, L. Zhao, and D. Metaxas, "CR-GCN: Channel-wise Recurrent Graph Convolutional Networks for Skeleton-based Action Recognition," in Proc. ECCV, pp. 1–17, 2020.

[6]. A. Nishida, H. Ogata, and Y. Kuniyoshi, "Robust Eye-Gaze Estimation in Unconstrained Settings Using Attention-Guided CNNs," IEEE Transactions on Human-Machine Systems, vol. 50, no. 4, pp. 317–327, 2020.

[7]. T. Luong, M. Hossain, and A. Sim, "DeepGaze: A Real-Time Gaze Estimation System for Assistive Communication," in Proc. ACM ICMI, pp. 22–30, 2021.

[8]. P. Patel and S. Kumar, "EyeSpeak: Low-Cost Eye Tracking for the Speech Impaired Using Embedded Vision," Sensors, vol. 21, no. 5, pp. 1663, 2021.

[9]. R. Sharma and P. Narang, "Gaze-Based HCI for Non-Verbal Users Using Hybrid CNN-LSTM," in Proc. IEEE International Conference on Smart Computing, pp. 1–6,

2022.
[10]. L. Zhang, Y. Feng, and K. Nakayama, "Lightweight Eye Tracker Using Mobile Front Camera and Vision Transformers," in Proc. CVPR Workshops, pp. 45–53, 2023.
[11]. N. Kapoor and V. K. Ojha, "Gaze-Controlled Systems Using Deep Learning for Assistive Technology: A Review," Multimedia Tools and Applications, vol. 81, pp. 17845–17862, 2022.
[12]. D. Park, M. Kim, and J. Ryu, "GazeNet: A Lightweight Eye Gaze Estimator for Real-Time User Interface Control," Applied Sciences, vol. 11, no. 14, 2021.
[13]. R. Behera, S. Ghosh, and A. Dash, "Real-Time Eye Tracking Using Embedded Platforms for Accessibility," IEEE Access, vol. 10, pp. 12562–12574, 2022.
[14]. M. Al-Husban and K. Al-Ja'afreh, "Web-Based Eye Tracking for Communication Aid Using CNN and WebRTC," Computers, vol. 10, no. 3, 2021.
[15]. J. Kim and Y. Jang, "Deep Learning-Based Eye Typing System for Locked-in Patients," in Proc. EMBC, pp. 4202–4205, 2020.
[16]. A. Mohan, P. Nair, and R. Sheth, "Towards Gaze-Based Text Entry Using Multi-Stage CNN-RNN Pipeline," Pattern Recognition Letters, vol. 148, pp. 23–31, 2021.
[17]. T. Singh and M. Singh, "Gaze-Based Virtual Keyboard with Deep Learning," in Proc. ICCCIS, pp. 1–6, 2022.
[18]. S. Mandal, A. Ray, and S. Ghosh, "Real-Time Blink and Gaze Detection for Assistive Communication," IEEE Transactions on Biomedical Circuits and Systems, vol. 15, no. 3, pp. 568–579, 2021.
[19]. Y. Iwata, T. Okada, and N. Maruyama, "Low-Latency Gaze Estimation Using Mobile Devices for Real-Time AAC," Sensors, vol. 21, no. 22, pp. 7712, 2021.
[20]. C. Wang, B. Liu, and X. Han, "Mobile Eye-Tracking for Real-Time Cognitive Load Estimation," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 30, pp. 129–137, 2022.
[21]. R. Puri, A. Desai, and A. Saxena, "Gaze-Assisted Communication Interface Using CNN and Eye-Blink Sensing," Journal of Assistive Technologies, vol. 17, no. 1, pp. 55–66, 2023.
[22]. V. Subramanian and S. D. Roy, "Attention-Based Gaze Estimation Framework for Remote Communication Aids," in Proc. IJCAI, pp. 2949–2956, 2022.
[23]. A. Tripathi and A. Maheshwari, "Optimizing Eye-Based Typing Systems Using Transformer Architectures," Computer Vision and Image Understanding, vol. 228, 103569, 2023.
[24]. M. Singh and A. Rai, "Real-Time Eye Gaze and Blink Detection for Virtual Keyboard Control," Human- Centric Computing and Information Sciences, vol. 13, no. 1, pp. 1–18, 2023.
[25]. N. Goel and H. Gupta, "EyeSpeak++: A Deep Learning-Powered Assistive Communication System Using Webcam Input," in Proc. EUSIPCO, 2022.
[26]. A. Rezaei and H. Aghajan, "Gaze-Driven Interaction Using Head Pose and Pupil Estimation on Edge Devices," Neurocomputing, vol. 493, pp. 81–91, 2022.
[27]. J. Patel and M. Thomas, "A Real-Time Gaze Interface for ALS Patients Using Embedded Vision and TTS," Assistive Technology, vol. 35, no. 1, pp. 44–52, 2023.
[28]. Z. Li and F. Wu, "End-to-End Deep Gaze-Based Intent Recognition in Assistive UI Systems," Journal of Vision Research, vol. 192, 107070, 2022.
[29]. S. Rahman and M. Shah, "Smart Eye: Vision-Based Eye Movement Analysis for Silent Communication," in Proc. ACM CHI, pp. 335–345, 2023.

[30]. B. Das and N. Verma, "Edge-AI Powered Gaze-Based Virtual Keyboard for Mobility-Impaired Users," IEEE Internet of Things Journal, vol. 11, no. 2, pp. 1987–19