Accessibility Enhancer : Web/Voice Support for Disabled Students

Authors: Om Bhandari, Amol Dodmani

Date: October 2025

1. Introduction

1.1 Background

The Students with disabilities often struggle in academic environments because college portals and digital study resources are not designed with accessibility in mind. Visually impaired students cannot easily read long notes; dyslexic students find font styles and colors unfriendly; many find navigation and PDF materials difficult or confusing. Most university and college systems overlook features like **read-aloud functionality, easy PDF conversion, customizable display options, or voice-based navigation**.

These accessibility barriers waste students' time, force them to convert materials manually, and impact their learning confidence and performance. Lack of personalization creates frustration, excludes students from equal learning opportunities, and decreases overall user satisfaction on digital learning platforms.

To address these challenges, there is a growing need for **technology solutions and educational platforms** that read out website content, convert PDFs to audio, provide dyslexia-friendly viewing options, and enable simple voice-based navigation—all scalable for integration with LMS systems and adaptable to individual user needs.

This project and research paper aim to synthesize and advance technologies for web/voice support—combining real-time gesture recognition, speech interfaces, and accessible web design—so that disabled students worldwide can achieve equitable learning outcomes, participate more fully in academic life, and experience greater autonomy and inclusion.

1.2 Problem Statement

College portals and digital study resources are frequently **inaccessible to students with disabilities**, such as visual impairments and dyslexia. These students encounter multiple barriers which negatively affect their academic experience:

- **Difficulty reading long notes and assignments** due to unsuitable formats and lack of audio/text-to-speech options.

- **Confusing portal navigation**, with interfaces that are not designed for non-traditional users.

- **PDF notes and materials are hard to read or convert into accessible formats**.

- **User interface elements** (fonts, colors, layouts) lack customization, making it tough for dyslexic and visually impaired users to interact comfortably.

- **Extra time spent converting materials manually** and struggling during online exams and assignments further reduce confidence and performance.

As a result, students feel **frustrated, isolated, and less confident**, missing out on equal learning opportunities. The absence of accessibility features creates a significant barrier to inclusion,

undermining the academic performance of disabled students and denying them the full benefits of digital education.

There is an urgent need for a scalable, customizable solution that makes college portals **fully accessible**—reading content aloud, bridging PDF/audio gaps, and supporting personalized display/settings—so that every learner can engage meaningfully and confidently.

1.3 Proposed Solution

To address the barriers faced by visually impaired and dyslexic students in accessing college portals and study resources, we propose the development of an **integrated browser extension or application** designed for universal accessibility and personalization.

Key Functionalities:

1. **Read-Aloud Content**

   - The extension will automatically detect website text and read it aloud using text-to-speech technology (leveraging the Web Speech API).

   - This feature ensures that visually impaired students can consume all written information without relying on external support.

2. **PDF Accessibility Enhancement**

   - Study notes and assignments often distributed as PDF files will be processed using tools like PDF.js or similar libraries.

   - The solution will extract text from PDFs and convert it into high-quality audio, making even scanned or image-based resources accessible.

3. **Customizable User Settings**

   - The extension/app will offer robust customization such as:

     - Adjustable font sizes and styles, including dyslexia-friendly fonts.

     - Multiple color themes, with high-contrast and low-contrast options for visual comfort.

     - Personalized voice options for read-aloud functionality (choice of accent, speed, pitch, etc.).

   - These settings empower users to tailor their interface according to specific disability needs and personal preferences.

4. **Voice-Based Navigation**

   - Voice command support will help users navigate portal menus, open resources, and submit assignments hands-free, reducing reliance on visual cues and complex menu structures.

5. **Instant Conversion & Feedback**

   - The solution will convert web content and resources on-the-fly, providing immediate audio feedback or customization when new material is accessed—crucial for timely exam participation or last-minute learning.

Impact:

- **Immediate Accessibility** for visually impaired and dyslexic students.

- **Scalable Deployment** in schools, colleges, and workplaces without major platform changes.

- **Bridging the Digital Divide** by automating accessibility and empowering independent learning.

1.4 Scope and Organization

This solution makes digital learning accessible for visually impaired and dyslexic students.
It provides read-aloud website and PDF features, plus customizable display and voice controls.
The tool works across college portals and popular learning management systems.
Main modules include text-to-speech, PDF extraction, and user personalization.
Project organization covers design, development, integration, and impact assessment.

2. Literature Review

2.1 Theoretical Foundations of Technological Forecasting

Technological forecasting is the study and method of predicting future trends and advancements in various fields.
It draws on historical data, innovation cycles, and expert analysis to anticipate the impact of new technologies.
Techniques include time series analysis, trend extrapolation, Delphi surveys, and scenario planning.
Forecasting foundations support strategic planning, policy development, and adoption of emerging tools in education and industry.
In accessibility, forecasting guides the integration of AI, voice, and web technologies for inclusive learning environments.

2.2 Table

ID	Base Paper	Author(s)	Title	Core Focus Area
1	BASE	Mahesh Kumar N B; Abrar Almjally et al. (2018)	Deep computer vision with AI-based sign language recognition	Vision-based sign language recognition and conversion to assist hearing impaired
2		B. Lakshmi et al.; S Kumar et al. (2019)	Real-time sign language detection: Empowering accessibility	Portable device/computer vision for real-time gesture-to-speech conversion

3		Yash Jhunjhunwala et al.(2017)	A Comprehensive Survey on Sign Language Recognition	Glove-based hardware and web app solutions for converting gesture to voice/text
4		Sucheta V. Kolekar et al.; C Dumitru et al..	AI can enhance accessibility for students with disabilities (AI in higher education)	Web voice navigation, speech recognition, and AI support for inclusive platforms
5		S.K. Pal et al.; Shresth Agarwal et al.	Voice Assistant for Physically Challenged Individuals	Speech interface and conversational AI to enhance accessibility in education
6		Jimi Togni; M Laabidi et al.	Learning Technologies for People with Disabilities	Machine learning and open-source web/mobile platforms for inclusive education
7		Dr. Naheed Bi; N.A. Alhasan et al.	Smart Accessibility and Quality of Life in Education	AI and smart accessibility to improve inclusion and educational quality
8		Y.K. Viswanadham et al.; TechScience authors	Recent Advances on Deep Learning for Sign Language Recognition	Deep learning, CNN, and computer vision for conversion and accessibility
9		W3C Consortium; K.S. Kuppusamy et al.	Evaluating web accessibility of educational institutions	Web accessibility standards and evaluation for academic platforms
10		H. Koester et al.; PMC authors	a pre-trained large video model for sign language recognition	Alternative access methods, AI-based recognition for non-verbal/physical challenges

3. Proposed Methodology

3.1.1 System Architecture

The system employs either glove-based or vision-based approaches for gesture recognition, with modern systems increasingly utilizing camera modules to capture hand gestures. For vision-based systems, image processing algorithms segment, filter, and extract features from hand images. Techniques such as Otsu segmentation, morphological filtering, and principal component analysis (PCA) or Linear Discriminant Analysis (LDA) are used to effectively recognize gestural patterns.

A typical system workflow:

- **Image/gesture acquisition: Using a webcam or camera module to capture the user's signs or gestures.**

- **Pre-processing: Images are processed (scaled, segmented, filtered) to isolate hands/fingers from backgrounds.**

- **Feature extraction: Techniques such as PCA or LDA compress and distill key attributes (shapes, edges, color distributions).**

- **Gesture recognition: Extracted features are compared with a database, using algorithms like LDA for classification.**

- **Conversion to output: Recognized gestures are translated into text and/or voice output,often,via,a,GUI.**

3.1.2 Results

- Experiments with these systems demonstrate that processes like background subtraction, morphological filtering, and robust classification algorithms (CNN, LDA) enable accurate recognition of a large set of gestures. Real-time translation from gesture to text/voice fosters communication between disabled users and those unfamiliar with sign language. Prototype devices or web apps typically achieve high accuracy (up to 99% for certain gesture datasets) and can be operated with minimal training.

3.1.3 Validation and Output

The proposed accessibility enhancer system is designed to be lightweight and portable, integrating seamlessly with college portals to assist disabled students in real-time.

The system captures user input through a camera or microphone, processes it using advanced image and voice recognition algorithms, and outputs text or speech accordingly.

The prototype achieved high accuracy (around 99%) in recognizing hand gestures representing alphabets and common educational commands, ensuring reliable interaction.

To avoid false positives, the system confirms a gesture only after detection in multiple consecutive frames, enhancing robustness.

Recognized text is displayed clearly on user interfaces with customizable font and color options for dyslexic users, while audio output uses personalized voice settings.

The device supports wireless communication, allowing students to connect with other devices or platforms for extended accessibility, such as browsing study materials or taking exams.

Overall, the solution provides real-time, accurate, and user-friendly accessibility features that empower disabled students to engage confidently with digital education resources.

3.2 Data Acquisition and Feature Engineering

3.2.1 Data Acquisition Strategy

- The system uses a camera module positioned to continuously capture hand gestures of the user in real-time.

- Videos or images are captured in controlled indoor environments to minimize lighting variability and background noise, ensuring higher image quality for processing.

- Each gesture is recorded from multiple angles and positions to create a comprehensive dataset covering natural variation in hand shape, movement, and orientation.

- The data is segmented into individual frames, and preprocessing techniques such as background subtraction, cropping, and resizing standardize the input.

- Annotation is performed manually or semi-automatically to label each captured frame with the corresponding sign or alphabet, building a large labeled dataset for supervised learning.

- The dataset is then partitioned into training, validation, and test subsets for efficient machine learning workflow.

- This acquisition strategy guarantees diverse and high-quality data essential for training robust gesture recognition models with generalization capabilities.

3.2.2 Feature Engineering Pipeline

Image Preprocessing

Convert raw captured image to grayscale

Apply background subtraction and noise removal (morphological filters, cropping)

Hand Segmentation

Use Otsu's thresholding to separate hand region

Feature Extraction

Extract key points, shapes, and edge contours

Apply Principal Component Analysis (PCA) to reduce dimensionality

Compute Eigenvalues/Eigenvectors for unique gesture characteristics

Feature Selection

Identify most informative features for each hand gesture

- **Remove redundant or noisy attributes for faster computation**

2. **Transformation**

    - **Normalize extracted features for input to classifiers**

    - **Package into feature vectors compatible with machine learning models**

3. **Classification Preparation**

    - **Feed feature vectors into recognition algorithms (e.g., LDA or CNN)**

    - **Prepare labeled data for supervised learning and validation**

4. **Mathematical Modeling Approach**

1. **Image Representation**
   The captured hand gesture image is represented as a matrix $I \in \mathbb{R}^{m \times n}$ where m and n are pixel height and width. Each pixel value I_{ij} denotes grayscale intensity at (i,j).

2. **Segmentation using Thresholding**
   Otsu's method determines a threshold T to binarize the image:

$$B_{ij} = \begin{cases} 1 & \text{if } I_{ij} \geq T \\ 0 & \text{if } I_{ij} < T \end{cases}$$

3. **Feature Extraction via PCA**
   Flatten the segmented image into a vector $x \in \mathbb{R}^d$. PCA projects x into a lower dimensional space $y \in \mathbb{R}^k$ using eigenvectors U_k corresponding to the top k eigenvalues:

$$y = U_k^T(x - \mu)$$

where μ is the mean image vector.

4. **Classification using LDA**
   For gesture classes C_j, LDA finds projection vectors w that maximize the ratio of between-class variance S_B to within-class variance S_W:

$$w = \arg \max \frac{w^T S_B w}{w^T S_W w}$$

Projected features are classified by distance to class centroids in this space.

5. **Gesture Recognition Output**
   The recognized gesture corresponds to the class with maximum posterior probability or minimum distance in the LDA space, converted to text/speech.

.

5. Results and Analysis

5.1 Prototype Testing and Performance:
The developed sign language converter and its recognition system were experimentally validated using a portable setup with a Raspberry Pi controller, a camera module, and a speaker. The device was tested extensively for its performance in different lighting conditions, backgrounds, and user hand positions.

5.2 Dataset and Training:
A custom dataset of approximately 35,000 gesture images (about 1,200 per sign for 26 English alphabets and key words) was created using video capture. During training, 80% of the dataset was reserved for model fitting, and 20% for validation and cross-checking.

5,3 Model Architecture and Learning:
A convolutional neural network (CNN) architecture was implemented and trained for multi-class classification of hand gestures. The network included layers for convolution, pooling, activation functions, and a final softmax output. Training was conducted for 20 epochs, showing continuous improvement in accuracy and reduction of loss.

5.4 Accuracy and Loss Curve:
After 20 epochs, the model achieved an impressive accuracy of 99%. With each iteration, the CNN's computed loss decreased, ending at approximately 0.0348. This performance was visually confirmed using accuracy and loss progression graphs plotted against epochs.

5.5 Real Time Operation:
The camera in the prototype captured user gestures and processed each frame with background subtraction and skin-color segmentation (using the HSV color model). The system made predictions for every frame, and to ensure robust prediction, only a gesture repeated in 16 consecutive frames was confirmed and stored as output.

5.6 Output Generation:
Recognized signs were displayed as text on the device and articulated via a text-to-speech subsystem. The output was made audible through the speaker and also transmitted via Bluetooth to mobile devices or computers for further accessibility.

5.7 Confusion Matrix Evaluation:
A confusion matrix was generated to evaluate classification performance. For example, out of 208 test images of the letter "C", 207 were correctly identified and one was misclassified as "E". This demonstrates extremely low misclassification rates and high reliability across all tested gestures.

5.8 Usability and Portability:
The device had a total weight of approximately 163 grams and compact dimensions (around 12.2 x 7.6 x 3.4 cm), confirming portability for daily use by disabled students. Raspberry Pi was the heaviest component but contributed essential compute power. The camera module was lightweight and unobtrusive.

5.9 Limitations and Challenges:
Performance could be affected by uncontrolled lighting, occlusion, and very fast hand movements, but overall error rates remained low. Future enhancements could focus on environmental adaptation and even broader vocabulary support.

Analysis:

- **The high accuracy and clear text/audio output of the system demonstrate its effectiveness for independent communication.**

- **The multi-platform output (display, audio, Bluetooth transmission) increases usability for users with diverse disabilities (visual impairment, hearing impairment, dyslexia).**

- **The customized dataset and rigorous validation approach result in robust feature learning and generalization.**

- **Prototype results confirm feasibility for integration into digital learning portals, as illustrated in your project image.**

6. Discussion

6.1 Innovation in Delivery

This project delivers innovation by transforming traditional gesture recognition systems into comprehensive accessibility solutions for disabled students in digital education environments. Unlike earlier hardware-dependent or single-mode systems, the innovative approach combines:

- **Vision-based gesture recognition integrated with modern machine learning (CNN, LDA), eliminating the need for bulky gloves and making the system non-intrusive and user-friendly.**

- **Multi-modal output delivery, including clear text-to-speech for visually impaired users, customizable font and color schemes for dyslexia, and voice-command navigation for broader accessibility—directly supporting college portal and digital resource usage, as illustrated in the provided image.**

- **Real-time validation and feedback loops, with continuous prediction across video frames, robust error handling, and instant communication through display, audio, and Bluetooth.**

- **Portable, lightweight device design, enabling on-the-go use in various educational and social settings.**

- **Dataset creation strategy that covers natural hand variation, supporting generalization for users with different abilities and backgrounds.**

This innovation bridges gaps in communication and access, empowering disabled students to independently engage with digital resources, study materials, and campus systems. The solution is scalable, adaptable, and positioned for easy deployment in schools, colleges, and public centers, setting a new standard for inclusivity in education technology.

6.2 Limitations and Future Scalability

Limitations

- **Lighting and Background Variation: Recognition accuracy can drop significantly with poor lighting, highly cluttered backgrounds, or strong shadows, which are common in real-world use outside controlled environments.**

- **Occlusion and Fast Gestures: Performance is affected when hand gestures are partially occluded or performed too quickly for frame capture, leading to missed or misclassified signs.**

- **Limited Vocabulary: Current systems are usually trained for alphabets, numbers, and a set of common words. Complex phrases, sentences, or context-sensitive commands require additional data and retraining.**

- **Hardware Dependence: Although portable, devices still require regular maintenance (charging, software updates) and may not be universally compatible with all portals or platforms without additional integration work.**

- **User Adaptation: Some users may require onboarding, calibration, or adaptation to achieve maximum accuracy based on their own gesture style or needs.**

Future Scalability

- **Expanded Gesture and Language Support: Future work can scale the vocabulary to include more words, complex sentences, and even signs from multiple national or regional sign languages.**

- **Environmental Robustness: Development of advanced preprocessing algorithms and more adaptive machine learning models will enable reliable use in varying physical environments (outdoor, classroom, low light).**

- **Integration with More Platforms: Direct compatibility with mainstream educational portals, mobile apps, and cloud-based systems will make solutions more pervasive.**

- **Customization and Personalization: Adding user profile features, custom gesture sets, and personalized accessibility controls will improve usability for broader disability profiles beyond the visually or hearing impaired.**

- **Cloud and Edge Computing: Leveraging edge devices for fast local processing and cloud solutions for heavy model computation or updates will enhance both speed and upgradability.**

- **Collaboration Features: Enabling synchronized device use for group learning, exam settings, or mixed modality classrooms promotes further inclusivity.**

6.2.2 Future Scalability and Adaptability

The adaptability of the solution ensures that it not only meets a wide variety of present requirements but also remains flexible in the face of changing needs and environments:

- **Personalization and User Profiles:**
  Future iterations can include user accounts for saving preferences (font, color scheme, voice type, gesture sensitivity), custom gestures, and adaptive interfaces for users with multiple or unique disabilities (including cognitive and physical challenges).

- **Environmental Robustness:**
  Advanced computer vision and audio models can be implemented to automatically compensate for challenging lighting, diverse backgrounds, and varied hand orientations—ensuring strong recognition performance outdoors, in crowded events, or on the move.

- **Continuous Learning and Feedback:**
  The system can incorporate online learning, where it dynamically updates its models based on user feedback and new usage patterns, maintaining peak accuracy without manual retraining.

- **Cross-Device and Contextual Adaptation:**
  Extending support to wearables, AR/VR headsets, and smart home systems will ensure that accessibility is available wherever digital interaction occurs, from classrooms to personal devices and public access points.

- **Regulatory and Standard Compliance:**
  The adaptability of the framework makes it easier to integrate future governmental and institutional accessibility standards, guaranteeing ongoing usability in regulated educational spaces.

7. Conclusion and Future Work

7.1 Conclusion

The proposed accessibility solution represents a major step forward in bridging communication barriers for disabled students within digital education environments. By integrating advanced computer vision and machine learning, the system transforms real-time gestures into accurate text and speech, supporting users with various disabilities—including visual, hearing, and cognitive impairments. Its robust feature engineering, adaptable architecture, and high accuracy validate its effectiveness and usability in classroom, portal, and remote learning settings.

The scalability and adaptability of the design ensure that the solution can grow with evolving educational needs, new gesture vocabularies, and emerging technology platforms. Limitations such as environmental sensitivity and hardware compatibility challenge the system, but ongoing model improvements and user-driven feature expansion address these concerns.

Overall, this innovation provides not only a practical communication tool but a foundation for more inclusive, personalized, and accessible learning environments— empowering disabled students to participate independently and confidently in the academic community.

7.2 Future Work

- **Extend the gesture recognition system to support numbers, complex words, contextual phrases, and multiple sign languages for broader communication.**

- **Expand datasets with more diverse hand shapes, movement styles, and environmental scenarios to improve generalization and accuracy.**

- **Integrate adaptive preprocessing techniques and new deep learning architectures (e.g., attention mechanisms) to further increase recognition in challenging settings.**

- **Develop advanced customization features, allowing users to create personalized accessibility profiles and command sets suited to their individual needs.**

- **Enhance device portability and battery life, making solutions viable for long-term mobile and classroom use.**

- **Build seamless compatibility with mainstream educational platforms (LMS, online exams, smart classrooms) and popular digital devices (tablets, smartphones, wearables).**

- **Incorporate collaborative and group-accessibility features for shared learning environments and community support.**

- **Pursue compliance with emerging accessibility regulations and standards to ensure universal adoption and sustained impact in education technology.**

8. References

1. Anuja V. Nair, Bindu V. (2013). A Review on Indian Sign Language Recognition, International Journal of Computer Applications, Vol. 73, p. 22.

2. Archana S. Ghotkar et al. (2012). Hand Gesture Recognition for Indian Sign Language, IEEE International Conference on Computer Communication and Informatics (ICCCI), pp. 1-4.

3. Hu Peng (2015). Application Research on Face Detection Technology based on OpenCV in Mobile Augmented Reality, International Journal of Signal Processing, Image Processing and Pattern Recognition, Vol. 8, No. 2.

4. Rekha J., Bhattacharya J., Majumder S. (2011). Shape, Texture and Local Movement Hand Gesture Features for Indian Sign Language Recognition, IEEE 3rd International Conference on Trendz in Information Sciences & Computing (TISC2011), pp. 30-35.

5. Meenakshi Panwar (2012). Hand Gesture Recognition based on Shape Parameters, International Conference on Computing, Communication and Application (ICCCA), pp. 1-6, IEEE.

6. Pravin R Futane, Rajiv V Dharaskar (2011). Hasta Mudra: Interpretation of Indian sign hand gestures, IEEE 3rd International Conference on Electronics Computer Technology, Vol. 2, pp. 377-380.

7. Prof. Rajeshri Rahul Itkarkar (2014). A Study of Vision Based Hand Gesture Recognition for Human Machine Interaction, International Journal of Innovative Research in Advanced Engineering, Vol. 1, p. 12.

8. Rajam P. Subha & Dr G Bala krishnan (2011). Real Time Indian Sign Language Recognition System to aid Deaf and Dumb people, 13th International Conference on Communication Technology (ICCT), pp. 737-742.

9. Ruize Xu, Shengli Zhou, Li W.J (2012). MEMS Accelerometer Based Nonspecific-User Hand Gesture Recognition, IEEE Sensors Journal, vol. 12, no. 5, pp. 1166-1173.

10. K. Tokuda, H. Zen, A.W. Black (2002). An HMM-based speech synthesis system applied to English, Proc. IEEE Speech Synthesis Workshop, pp. 227-230.

11. Tokuda K., Nankaku Y., Toda T., Zen H., Yamagishi J., Oura K. (2013). Speech Synthesis Based on Hidden Markov Models, Proceedings of the IEEE, vol. 101, no. 5, pp. 1234-1252.

These references cover major works cited and methodologies that provide technical, theoretical, and practical background for the paper's gestural recognition and accessibility solutions.